

GenCore version 5.1.7

Copyright (c) 1993 - 2006 Bioacceleration Ltd.

OM nucleic - nucleic search, using sw model

Run on: February 9, 2006, 14:43:53 ; Search time 4488.31 Seconds
(without alignments)
11410.963 Million cell updates/sec

Title: US-10-607-806-1-T9182_COPY_8700_9600

Perfect score: 901

Sequence: 1 tggctattataataaaat.....gagctgaggcaggagatc 901

Scoring table: IDENTITY_NUC

Gapop 10.0 , Gapext 1.0

Searched: 5893141 seqs, 28421725653 residues

Total number of hits satisfying chosen parameters: 11766282

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

GenEmbl.*

1: gb_ba.*

2: gb_in.*

3: gb_env.*

4: gb_on.*

5: gb_ov.*

6: gb_pat.*

7: gb_ph.*

8: gb_pr.*

9: gb_ro.*

10: gb_sts.*

11: gb_sy.*

12: gb_un.*

13: gb_vi.*

14: gb_vt.*

15: gb_pl.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	901	100.0	220384	14 AC078926	AC078926 Homo sapi
2	899.4	99.8	8368	8 AY438977	AY438977 Homo sapi
3	899.4	99.8	13612	6 AX377239	AX377239 Sequence
4	899.4	99.8	122302	8 AC003982	AC003982 Homo sapi
5	897.8	99.6	189729	14 AC073930	AC073930 Homo sapi
6	337.2	37.4	173950	8 AL390961	AL390961 Human DNA
7	337.2	37.4	174937	14 AC022245	AC022245 Homo sapi
8	337.2	37.4	179245	14 AL591164	AL591164 Homo sapi
9	337.2	37.4	266738	14 AL138896	AL138896 Homo sapi
10	324.8	36.0	100506	14 AP002020	AP002020 Homo sapi
11	324.8	36.0	118642	8 AC126283	AC126283 Homo sapi
12	324.2	36.0	161306	14 AC091056	AC091056 Homo sapi
13	324.2	36.0	162100	8 AC018868	AC018868 Homo sapi
14	324.2	36.0	180230	14 AC090783	AC090783 Homo sapi
15	323.8	35.9	154592	8 AL591215	AL591215 Human DNA
16	323.8	35.9	173176	14 AC022857	AC022857 Homo sapi
17	322.8	35.8	176438	14 AC126761	AC126761 Homo sapi
18	322.8	35.8	181265	14 AC083994	AC083994 Homo sapi

19	322.8	35.8	184777	8 AC026749	AC026749 Homo sapi
20	322.8	35.7	89673	8 AC004389	AC004389 Homo sapi
21	321.8	35.7	179993	14 AC015879	AC015879 Homo sapi
22	321.2	35.6	158158	14 AC034304	AC034304 Homo sapi
23	320	35.5	155405	8 AC009812	AC009812 Homo sapi
24	319.8	35.5	146181	8 AP001269	AP001269 Homo sapi
25	319.6	35.5	123288	8 AL354712	AL354712 Human DNA
26	318.6	35.4	190626	14 AC062036	AC062036 Homo sapi
27	318.6	35.4	112646	8 AC010240	AC010240 Homo sapi
28	318	35.3	166258	14 AL451066	AL451066 Homo sapi
29	317.8	35.3	134341	8 AC099794	AC099794 Homo sapi
30	317.6	35.2	88546	8 AC008435	AC008435 Homo sapi
31	317.2	35.2	37490	8 AC026356	AC026356 Homo sapi
32	317.2	35.2	220480	14 AC023973	AC023973 Homo sapi
33	317	35.2	167478	14 AC011148	AC011148 Homo sapi
34	317	35.2	176539	8 AC074213	AC074213 Homo sapi
35	317	35.2	182103	14 AC016998	AC016998 Homo sapi
36	316.8	35.2	167863	14 ALS90651	ALS90651 Homo sapi
37	316.6	35.1	171951	8 CNS05TDA	AL356020 Human chr
38	316.4	35.1	133512	8 AL590714	AL590714 Human DNA
39	315.6	35.0	157579	8 AC018639	AC018639 Human chr
40	315.6	35.0	159451	8 AC012618	AC012618 Homo sapi
41	315.6	35.0	139014	14 AC073849	AC073849 Homo sapi
42	315.6	35.0	224573	8 AC008758	AC008758 Homo sapi
43	315.2	35.0	97916	8 HS408N23	Z98048 Human DNA s
44	315.2	35.0	280708	14 AL161641	AL161641 Homo sapi
45	314.8	34.9	161256	8 AC087855	AC087855 Homo sapi

ALIGNMENTS

RESULT 1
AC078926/c
LOCUS Homo sapiens chromosome 12 clone RP11-836M11, WORKING DRAFT
DEFINITION 220384 bp DNA linear HTG 26-MAR-2002
SEQUENCE, 11 unordered pieces.
AC078926
AC078926.23 GI:19718703
HTG; HTGS PHASE1; HTGS_DRAFT; HTGS_FULLTOP.
KEYWORDS Homo sapiens (human)
SOURCE Homo sapiens
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini; Hominidae; Homo.
REFERENCE
1 (bases 1 to 220384)
Muzny,D.M., Adams,C., Adio-Oduola,B., Ali-osman,F.R., Allen,C., Alsbrooks,S.L., Amaratunga,H.C., Are,J.R., Ayele,M., Banks,T., Barbara,J., Benton,J., Bimage,K., Blankenburg,K., Bonnin,D., Bouck,J., Bowie,S., Brieva,M., Brown,E., Brown,M., Bryant,N.P., Buhay,C., Burch,P., Burkett,C., Burrell,K.L., Byrd,N.C., Carron,T.F., Carter,M., Cavazos,S.R., Chacko,J., Chavez,D., Chen,G., Chen,R., Chen,Z., Chowdhry,I., Christopoulos,C., Cleveland,C.D., Cox,C., Coyle,M.D., Dathorne,S.R., David,R., Davila,M.L., Davis,C., Davy-Carroll,L., Dederich,D.A., Delaney,K.R., Delgado,O., Denn,A.L., Ding,Y., Dinh,H.H., Douthwaite,K.J., Draper,H., Dugan-Rocha,S., Durbin,K.J., Earnhart,C., Edgar,D., Edwards,C.C., Elhaj,C., Escotto,M., Falls,T., Ferraguto,D., Flagg,N., Ford,J., Foster,P., Frantz,P., Gabisi,A., Gao,J., Garcia,A., Garner,T., Garza,N., Gill,R., Gorrell,J.H., Guevara,W., Gunaratne,P., Hale,S., Hamilton,K., Harris,C., Harris,K., Hart,M., Havlak,P., Hawes,A., Hernandez,J., Hernandez,O., Hodgeson,A., Hughes,M., Holloway,C., Hollins,B., Homsi,F., Howard,S., Huber,J., Hulyk,S., Hume,J., Jackson,L.E., Jacobson,B., Jia,Y., Johnson,R., Jolivet,S., Joudah,S., Karlsson,E., Kelly,S., Khan,U., King,L., Korvah,J., Kovar,C., Kratovic,J., Kureshi,A., Landry,N., Leal,B., Lewis,L.C., Lewis,L., Li,J., Li,Z., Lichtarge,O., Lien,C., Liu,J., Liu,W., Loulssed,H., Lozadó,R.J., Lu,X., Lucier,A., Lucier,R., Luna,R., Ma,J., Maheshwari,M., Mapua,P., Martin,R., Martindale,A., Martinez,E., Massy,E., Mawhney,E., McLeod,M.P., Meador,M., Mei,G., Metzger,M., Miner,G., Miner,Z., Mitchell,K., Mohabbat,K., Morgan,M., Morris,S., Moser,M., Neal,D., Newton,J., Newton,N., Nguyen,A., Nguyen,N.,

Nguyen, N., Nickerson, E., Nwokenkwo, S., Oguh, M., Okwuonu, G., Orghueme, N., Ovielo, R., Pace, A., Payton, B., Peery, J., Perez, L., Peters, L., Pickens, R., Primus, E., Pu, L., Quiles, M., Ren, Y., Rivers, M., Rojas, A., Rojokan, I., Rolfe, M., Ruiz, S., Savery, G., Scherer, S., Scott, G., Shen, H., Shoohtari, N., Sisson, I., Sodergren, E., Sonaike, T., Sparks, A., Stanley, H., Stone, H., Sutton, A., Svatek, A., Tabor, P., Tamerisa, A., Tamerisa, K., Tang, H., Tansey, J., Taylor, C., Taylor, T., Telford, B., Thomas, N., Thomas, S., Usmani, K., Vaquez, L., Vera, V., Villalon, D., Vinson, R., Wang, Q., Wang, S., Ward-Moore, S., Warren, R., Washington, C., Watlington, S., Williams, G., Williamson, A., Wleczyk, R., Wooden, S., Worley, K., Wu, C., Wu, Y., Wu, Y. F., Zhou, J., Zorrilla, S., Nelson, D., Weinstock, G., and Gibbs, R.

Direct Submission
Unpublished
2 (bases 1 to 220384)
Worley, K.C.
Direct Submission
Submitted (11-AUG-2000) Human Genome Sequencing Center, Department of Molecular and Human Genetics, Baylor College of Medicine, One Baylor Plaza, Houston, TX 77030, USA
3 (bases 1 to 220384)
Worley, K.C.
Direct Submission
Submitted (26-MAR-2002) Human Genome Sequencing Center, Department of Molecular and Human Genetics, Baylor College of Medicine, One Baylor Plaza, Houston, TX 77030, USA
On Mar 26, 2002 this sequence version replaced gi:18449664.

----- Genome Center
Center: Baylor College of Medicine
Center code: BCM
Web site: <http://www.hgsc.bcm.tmc.edu/>
Contact: hgsc-help@bcm.tmc.edu
----- Project Information
Center project name: HCAH
Center clone name: RP11-836M11
----- Summary Statistics
Sequencing vector: M13;
Chemistry: Dye-primer Bodipy: 47% of reads
Assembly: Dye-terminator Big Dye: 53% of reads
Phrap: version 0.990329
Consensus quality: 234783 bases at least Q40
Consensus quality: 241821 bases at least Q20
Consensus quality: 245671 bases at least Q20
Estimated insert size: 219187; sum-of-contigs estimation
Quality coverage: 8.9x in Q20 bases; sum-of-contigs estimation

COMMENT

* NOTE: Estimated insert size may differ from sequence length (see http://www.hgsc.bcm.tmc.edu/docs/Genbank_draft_data.html).
* NOTE: This is a 'working draft' sequence. It currently consists of 11 contigs. The true order of the pieces is not known and their order in this sequence record is arbitrary. Gaps between the contigs are represented as runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence as soon as it is available and the accession number will be preserved.

* 1 2156: contig of 2156 bp in length
* 2157 2256: gap of unknown length
* 2257 7112: contig of 4856 bp in length
* 7113 7212: gap of unknown length
* 7213 13283: contig of 6071 bp in length
* 13284 13383: gap of unknown length
* 13384 20660: contig of 7277 bp in length
* 20661 20760: gap of unknown length
* 20761 34473: contig of 13713 bp in length
* 34474 34573: gap of unknown length
* 34574 44394: contig of 9821 bp in length
* 44395 44495: gap of unknown length
* 44496 57126: contig of 12632 bp in length
* 57127 57226: gap of unknown length
* 57227 80266: contig of 23040 bp in length

FEATURES	source	1..220384	80366: gap of unknown length
*	80367	109398: contig of 29032 bp in length	
*	80367	109398: gap of unknown length	
*	109399	149812: contig of 40314 bp in length	
*	109399	149812: gap of unknown length	
*	149813	220384: contig of 70472 bp in length	
*	149813	220384: gap of unknown length	
FEATURES	source	1..220384	80366: gap of unknown length
		/organism="Homo sapiens"	
		/mol_type="genomic DNA"	
		/db_xref="taxon:9606"	
		/chromosome="12"	
		/clone="RP11-836M11"	
		2157..2256	
gap		/estimated_length=unknown	
gap		7113..7212	
gap		/estimated_length=unknown	
gap		13284..13383	
gap		/estimated_length=unknown	
gap		20661..20760	
gap		/estimated_length=unknown	
gap		34474..34573	
gap		/estimated_length=unknown	
gap		44395..44494	
gap		/estimated_length=unknown	
gap		57127..57226	
gap		/estimated_length=unknown	
gap		80267..80366	
gap		/estimated_length=unknown	
gap		109399..109498	
gap		/estimated_length=unknown	
gap		149813..149912	
gap		/estimated_length=unknown	
ORIGIN		Query Match 100.0%; Score 901; DB 14; Length 220384;	
		Best Local Similarity 100.0%; Pred. No. 6.3e-172;	
		Matches 901; Conservative 0; Mismatches 0; Indels 0; Gaps 0;	
QY	1	TGTTTCATATTAATAAATAAATCACTGTTGGGACAGTCAAGCTGTAACACAGT	60
Db	69811	TGTTTCATATTAATAAATAAATCACTGTTGGGACAGTCAAGCTGTAACACAGT	69752
QY	61	ACTTTGGGAAGTCCCAAGGTGGGTGGATCACTTGGGTGAGAAGTTCCGACACAGCCTGGTC	120
Db	69751	ACTTTGGGAAGTCCCAAGGTGGGTGGATCACTTGGGTGAGAAGTTCCGACACAGCCTGGTC	69692
QY	121	AATATGGTGAACCCCTATCTCTACTATAAATAACAAAAATTAGCTGGGTGTAGTGTGAT	180
Db	69691	AATATGGTGAACCCCTATCTCTACTATAAATAACAAAAATTAGCTGGGTGTAGTGTGAT	69632
QY	181	GCCTGTAGTCCAGCTACTCGGGAGGCTGAGGCAAGAGAAATGCTTGAACCTGGGAGGCA	240
Db	69631	GCCTGTAGTCCAGCTACTCGGGAGGCTGAGGCAAGAGAAATGCTTGAACCTGGGAGGCA	69572
QY	241	GAGGTTCAGTTCAGCCGAGATCCACACTCCAGCTCCAGCTGGGCGACACAGCCGAGCT	300
Db	69571	GAGGTTCAGTTCAGCCGAGATCCACACTCCAGCTCCAGCTGGGCGACACAGCCGAGCT	69512
QY	301	CTATCTCAAAAAATAAATAAATAAATAAAGGATCGGAGAGAAAACAAAACTAATAAGAT	360
Db	69511	CTATCTCAAAAAATAAATAAATAAATAAAGGATCGGAGAGAAAACAAAACTAATAAGAT	69452
QY	361	TCCTGAAGGTAAAGCAGAGATACGTAAATTAATGTAATAAAGTTAAATGCAATTTAACT	420
Db	69451	TCCTGAAGGTAAAGCAGAGATACGTAAATTAATGTAATAAAGTTAAATGCAATTTAACT	69392
QY	421	GTAATCTTATTTGTTTATTTGTTTATAAAGTAAACCAAGCCAAAGTAACTTCAAA	480
Db	69391	GTAATCTTATTTGTTTATTTGTTTATAAAGTAAACCAAGCCAAAGTAACTTCAAA	69332
QY	481	ACTCTACATAAATAATCTATTTATGGAAGTGGAGGATCTATAATCTCTACCTACCAAGA	540

Db 69331 ACTCTACATAAATATCTATTATGGAAAGTGGAGGCACTATAATCTTACTACCCAAAGA 69272

QY 541 TAACCAAGTTACATAATTCCTCCAGATTTTGGGGCATACACTAGCTTTTTTATTTGGGAA 600

Db 69271 TAACCAAGTTACATAATTCCTCCAGATTTTGGGGCATACACTAGCTTTTTTATTTGGGAA 69212

QY 601 AATTTCCATGTCGAGGCATACCTAATTTTCTTAATGCTATGTAGTATTCATTAAAGG 660

Db 69211 AATTTCCATGTCGAGGCATACCTAATTTTCTTAATGCTATGTAGTATTCATTAAAGG 69152

QY 661 ATGTTCCATAATTTTAAATAACATGCTTTAAAGTAGAGAACTAGGTTGGGCATGGTGG 720

Db 69151 ATGTTCCATAATTTTAAATAACATGCTTTAAAGTAGAGAACTAGGTTGGGCATGGTGG 69092

QY 721 CTCACGCTGTATCCAGCACTTTGGAGCGCGAGCAAAATGATCACTTGAGGTCCCGA 780

Db 69091 CTCACGCTGTATCCAGCACTTTGGAGCGCGAGCAAAATGATCACTTGAGGTCCCGA 69032

QY 781 GTTTGAGACCACTGGACAAATGATGAACACACCTCTCTTAATAAAATACAAAATTA 840

Db 69031 GTTTGAGACCACTGGACAAATGATGAACACACCTCTCTTAATAAAATACAAAATTA 68972

QY 841 GCTGGGCATGTTGGCAAGCACTGTAGTCCAGCTACTCAGGAGTCTGAGGCAGAGTAT 900

Db 68971 GCTGGGCATGTTGGCAAGCACTGTAGTCCAGCTACTCAGGAGTCTGAGGCAGAGTAT 68912

QY 901 C 901

Db 68911 C 68911

RESULT 2

AY438977

LOCUS 8368 bp DNA linear PRI 29-OCT-2003

DEFINITION Homo sapiens phospholipase A2, group IB (pancreas) (PLA2G1B) gene, complete cds.

ACCESSION AY438977

VERSION AY438977.1 GI:37953284

KEYWORDS

SOURCE Homo sapiens (human)

ORGANISM Homo sapiens

REFERENCE 1 (bases 1 to 8368)

AUTHORS Rieder, M.J., Livingston, R.J., Daniels, M.R., Chung, M.-W., Miyamoto, K.E., Nguyen, C.P., Nguyen, D.A., Poel, C.L., Robertson, P.D., Schackwitz, W.S., Sherwood, J.K., Witrak, L.A. and Nickerson, D.A.

TITLE Direct Submission

JOURNAL Submitted (16-OCT-2003) Genome Sciences, University of Washington, 1705 NE Pacific, Seattle, WA 98195, USA

COMMENT To cite this work please use: NIEHS-SNPs, Environmental Genome Project, NIEHS ES15478, Department of Genome Sciences, Seattle, WA (URL: <http://egp.gs.washington.edu>).

FEATURES

source

1..8368

/organism="Homo sapiens"

/mol_type="genomic DNA"

/db_xref="taxon:9606"

101..257

/rpt_family="MIR"

repeat_region

205

/rpt_type=dispersed

variation

205

/frequency="0.01"

/replace="a"

variation

306

/frequency="0.01"

/replace="a"

repeat_region

326..538

/rpt_family="Alu"

/rpt_type=dispersed

misc_feature

354..496

/note="Region not scanned for variation"

gene

875..6548

/gene="PLA2G1B"

join(875..944,2644..2803,3603..3730,6347..6548)

/gene="PLA2G1B"

/product="phospholipase A2, group IB (pancreas)"

join(911..944,2644..2803,3603..3730,6347..6471)

/gene="PLA2G1B"

/codon_start=1

/product="phospholipase A2, group IB (pancreas)"

/protein_id="AA05441.1"

/db_xref="GI:37953285"

/translation="MKLLVLAALLTVAADSGISPRVQPRKIKCVIPGSDPFLEY

NYCYCGLGSGTGVDELDKCCQCHDNCYDQAKLPSCKFLDNPTHTTYSYSCSGS

AITCSSKNKECEAFICNDRNAATCFSKAPYNKAHKNLDTKKYQCS"

1174..1307

/rpt_family="Alu"

/rpt_type=dispersed

variation

1286

/gene="PLA2G1B"

/frequency="0.01"

/replace="g"

repeat_region

1319..1588

/rpt_family="Alu"

/rpt_type=dispersed

variation

1343

/gene="PLA2G1B"

/frequency="0.03"

/replace="a"

variation

1593

/gene="PLA2G1B"

/frequency="0.02"

/replace="t"

repeat_region

1605..1669

/rpt_family="L2"

/rpt_type=dispersed

misc_feature

1638..2188

/gene="PLA2G1B"

/note="Region not scanned for variation"

repeat_region

1670..1975

/rpt_family="Alu"

/rpt_type=dispersed

repeat_region

1976..2038

/rpt_family="L2"

/rpt_type=dispersed

repeat_region

2086..2165

/rpt_family="L2"

/rpt_type=dispersed

repeat_region

2194..2319

/rpt_family="Alu"

/rpt_type=dispersed

misc_feature

2319..2453

/gene="PLA2G1B"

/note="Region not scanned for variation"

variation

3027

/gene="PLA2G1B"

/frequency="0.25"

/replace="t"

repeat_region

3185..3431

/rpt_family="Alu"

/rpt_type=dispersed

variation

3262

/gene="PLA2G1B"

/frequency="0.01"

/replace="t"

variation

3321

/gene="PLA2G1B"

/frequency="0.01"

/replace="a"

variation

3396

/gene="PLA2G1B"

/frequency="0.03"

/replace="g"

variation

3630

variation	/gene="PLA2G1B" /frequency="0.10" /replace="c" 3702		variation	/frequency="0.01" /replace="a" 5949		
	/gene="PLA2G1B" /frequency="0.15" /replace="a" 3768			/gene="PLA2G1B" /replace="g" 6039		
	/gene="PLA2G1B" /frequency="0.01" /replace="t" 3818			/gene="PLA2G1B" /replace="a" 6082		
	/gene="PLA2G1B" /frequency="0.01" /replace="a" 3846.3964			/gene="PLA2G1B" /frequency="0.14" /replace="t" 6082		
repeat_region	/rpt_family="MIR" /rpt_type=dispersed 3977.4280			Query Match 99.8%; Score 899.4; DB 8; Length 8368; Best Local Similarity 99.9%; Pred. No. 2.1e-171; Matches 900; Conservative 0; Mismatches 1; Indels 0; Gaps 0;		
repeat_region	/rpt_family="Alu" /rpt_type=dispersed 4281.4346			Qy 1	TGTTTCATATTAAATTAATAAATACTCAGTTGGGCACAGTCAAGCTGTAACACACAGT 60	
repeat_region	/rpt_family="L2" /rpt_type=dispersed 4347.4634			Db 5074	TGTTTCATATTAAATTAATAAATACTCAGTTGGGCACAGTCAAGCTGTAACACACAGT 5133	
	/rpt_family="Alu" /rpt_type=dispersed 4386			Qy 61	ACTTTGGAAGTCCAAGGTGGGTGGATCACTTGAAGTGAGAAGTTTCGAGACCAGCCTGGTTC 120	
variation	/gene="PLA2G1B" /frequency="0.08" /replace="t" 4635.4723			Db 5134	ACTTTGGAAGTCCAAGGTGGGTGGATCACTTGAAGTGAGAAGTTTCGAGACCAGCCTGGTTC 5193	
	/rpt_family="L2" /rpt_type=dispersed 5013			Qy 121	AATATGGTGAACCCCTATCTCTACTAAATAACAAAAATTAGCTGGGTGTAGTATGCAT 180	
repeat_region	/gene="PLA2G1B" /frequency="0.10" /replace="a" 5103.5404			Db 5194	AATATGGTGAACCCCTATCTCTACTAAATAACAAAAATTAGCTGGGTGTAGTATGCAT 5253	
variation	/rpt_family="Alu" /rpt_type=dispersed 5133			Qy 181	GCCTGTAGTCCCAAGCTACTCGGGAGGCTCAGGCAAGAGAAATTGCTTGAACCTGGGAGGCA 240	
	/gene="PLA2G1B" /frequency="0.01" /replace="c" 5543			Db 5254	GCCTGTAGTCCCAAGCTACTCGGGAGGCTCAGGCAAGAGAAATTGCTTGAACCTGGGAGGCA 5313	
repeat_region	/gene="PLA2G1B" /frequency="0.01" /replace="c" 5555			Qy 241	GAGGTTGCAGTGCAGCCGAGATCCACCACTGCACCTCCAGCCTGGGCGACACAGCGAGACT 300	
variation	/gene="PLA2G1B" /frequency="0.01" /replace="t" 5577.5629			Db 5314	GAGGTTGCAGTGCAGCCGAGATCCACCACTGCACCTCCAGCCTGGGCGACACAGCGAGACT 5373	
	/rpt_family="L1" /rpt_type=dispersed 5580			Qy 301	CTATCTCAAAAAATAATAATAATAAAGGATCGGAGAGAAAAACAAACTAATAAGAT 360	
variation	/gene="PLA2G1B" /frequency="0.01" /replace="c" 5592			Db 5374	CTATCTCAAAAAATAATAATAATAAAGGATCGGAGAGAAAAACAAACTAATAAGAT 5433	
	/gene="PLA2G1B" /frequency="0.01" /replace="c" 5595			Qy 361	TCCTGAAGGTAAACAGAGATACGTAATATATATTAATAAAGTTTAAATGCACTTTAACT 420	
variation	/gene="PLA2G1B" /frequency="0.01" /replace="t" 5598			Db 5434	TCCTGAAGGTAAACAGAGATACGTAATATATATTAATAAAGTTTAAATGCACTTTAACT 5493	
variation	/gene="PLA2G1B" /frequency="0.01" /replace="t" 5600			Qy 421	GTAATCTTATGTTTATTTTGGTTTATAAAGTAAACAGCCAAAGTAATGCAACTCAA 480	
	/gene="PLA2G1B" /frequency="0.15" /replace="g" 5602			Db 5494	GTAATCTTATGTTTATTTTGGTTTATAAAGTAAACAGCCAAAGTAATGCAACTCAA 5553	
repeat_region	/rpt_family="L1" /rpt_type=dispersed 5675			Qy 481	ACTCTACATAAATATCTATTATGGAAGTGAAGGCATCTATATCTCTACTACCCAAAGA 540	
	/rpt_family="L1" /rpt_type=dispersed 5675			Db 5554	ACTCTACATAAATATCTATTATGGAAGTGAAGGCATCTATATCTCTACTACCCAAAGA 5613	
variation	/gene="PLA2G1B" /frequency="0.09" /replace="g" 5675			Qy 541	TAACCCAGTTACATATTCCTCCAGATTTTGGGGCATACTAGCTTTTTTATTTGGGAA 600	
	/gene="PLA2G1B" /frequency="0.09" /replace="g" 5675			Db 5614	TAACCCAGTTACATATTCCTCCAGATTTTGGGGCATACTAGCTTTTTTATTTGGGAA 5673	
variation	/gene="PLA2G1B" /frequency="0.09" /replace="g" 5675			Qy 601	AATTTCCATGTGAGGCATACCTAAATTTTCTAAATGTCTATGTAGTATTTCCATTAAAG 660	
	/gene="PLA2G1B" /frequency="0.09" /replace="g" 5675			Db 5674	AATTTCCATGTGAGGCATACCTAAATTTTCTAAATGTCTATGTAGTATTTCCATTAAAG 5733	
variation	/gene="PLA2G1B" /frequency="0.09" /replace="g" 5675			Qy 661	ATGTTCCATAATTTTAAATAACATACGCTTTAAAGTAGAGAAACTAGGTTGGGCATGGTGG 720	
repeat_region	/rpt_family="Alu" /rpt_type=dispersed 5929			Db 5734	ATGTTCCATAATTTTAAATAACATACGCTTTAAAGTAGAGAAACTAGGTTGGGCATGGTGG 5793	
	/rpt_family="Alu" /rpt_type=dispersed 5929			Qy 721	CTCAGCCCTGTATCCAGCAGCTTTGGGAGGCGGAGCAATCGATCACTTTGAGTCCCGGA 780	
variation	/gene="PLA2G1B" /frequency="0.09" /replace="g" 5929			Db 5794	CTCAGCCCTGTATCCAGCAGCTTTGGGAGGCGGAGCAATCGATCACTTTGAGTCCCGGA 5853	
	/gene="PLA2G1B" /frequency="0.09" /replace="g" 5929			Qy 781	GTTTGAACCCAGCCTGGACAAACATGATGAAACACCCCTCTCTAATAAAAAATACAAAAATTA 840	

Genome Sequencing Center
Department of Genetics
Washington University
St. Louis MO 63108, USA
<http://genome.wustl.edu/gsc>
<mailto:sapiens@watson.wustl.edu>

NOTICE: This sequence may not represent the entire insert of this clone. It may be shorter because we only sequence overlapping clone sections once, or longer because we provide a small overlap between neighboring data submissions.

This sequence was finished as follows unless otherwise noted:
all regions were double stranded, sequenced with an alternate chemistry, or covered by high quality data (i.e., phred quality >= 30); an attempt was made to resolve all sequencing problems, such as compressions and repeats; all regions were covered by sequence from more than one subclone; and the assembly was confirmed by restriction digest.

MAPPING INFORMATION:

This clone was originally isolated in the laboratory of Professor Graeme Bell, Howard Hughes Medical Institute and Departments of Biochemistry and Molecular Biology, and Medicine, The University of Chicago, Chicago, IL, USA. The clone was provided by the laboratory of Dr. Roger Cox at The Wellcome Trust Centre for Human Genetics, Oxford, UK. Some contig information was also obtained from Yamagata et al., Nature 384:455-8 (1996).

SOURCE INFORMATION:

This clone was derived from human PAC library RPCI-1, prepared by Pieter de Jong and coworkers at the Roswell Park Cancer Institute (<http://bacpac.med.buffalo.edu>) using the method described by Ioannou et al., Nature Genetics 6:84-9 (1994). The library is from one male donor.
The clone may be obtained either from Genome Systems, Inc. (<http://www.genomesystems.com>) or Research Genetics, Inc. (<http://www.resgen.com>); or from Pieter de Jong.
VECTOR: pCYPAC2

NEIGHBORING SEQUENCE INFORMATION:

The clone sequenced to the left is 278C19; the clone sequenced to the right is 15E1. Actual start of this clone is at base position 1 of 166H1; actual end is at 122302 of 166H1.

FEATURES

source
1..122302
Location/Qualifiers
/organism="Homo sapiens"
/mol_type="genomic DNA"
/db_xref="taxon:9606"
/chromosome="12"
/map="12q"
/clone="166H1"
/clone_lib="RPCI-1"
repeat_region
1..228
/rpt_family="Alu"
repeat_region
239..265
/rpt_family="AT_rich"
repeat_region
268..565
/rpt_family="Alu"
repeat_region
1001..1309
/rpt_family="Alu"
repeat_region
1440..1745
/rpt_family="Alu"
repeat_region
1909..1940
/rpt_family="Alu"
repeat_region
1955..1983
/rpt_family="MER1_type"
repeat_region
2063..2362
/rpt_family="AT_rich"
repeat_region
2516..2650
/rpt_family="Alu"
repeat_region
2674..2730
/rpt_family="L1"
repeat_region
2735..2788

repeat_region
2790..3077
/rpt_family="(GAAA)n"
repeat_region
3081..3139
/rpt_family="Alu"
repeat_region
3716..4020
/rpt_family="L1"
repeat_region
4196..4492
/rpt_family="Alu"
gene
4560..14034
/rpt_family="Alu"
CDS
/gene="WUGSC:H.166H1.1"
join(4560..5056,13418..13749,13882..14034)
/gene="WUGSC:H.166H1.1"
/note="unknown function; 60% similar to Z50177 (PID:g927403) (PID:g927402); H_166H1.1"
/codon_start=1
/evidence=not_experimental
/protein_id="AA895634.1"
/db_xref="GI:276966"
/translation="MKMSFALTFRSAGRWIANPSPQCSKASIGLFPASPPLDPEKV KEIQRTITLSKRLIVMTGAGISTESGIPDYRSEKVGLYARTDRPIQHGDFVRSAPIR QRYWRFVGFQFSSHOPNPAHWALSTWEKLGKLYLVTONVDALHTKAGSRRLTEL HGCMDRAYCSVFLGSRVLCDCGEGTPRGVJOERQVNLNPTWSAEHGLAPGDVPL LSEQVRSFQVTCVQCQGHLPDVPFVFGTVNPKDVFVHKRVKEADSLLVGSSLLQ VYSGYRILTAWKKLPAILINIGPTRSDDLACLKLNRCGELLPLIDPC"
repeat_region
5366..5655
/rpt_family="Alu"
repeat_region
5672..5798
/rpt_family="Alu"
repeat_region
5802..5834
/rpt_family="(TA)n"
repeat_region
5834..5857
/rpt_family="POLY_A"
repeat_region
5870..6184
/rpt_family="Alu"
repeat_region
6224..6268
/rpt_family="(CA)n"
repeat_region
6269..6558
/rpt_family="Alu"
repeat_region
6729..7029
/rpt_family="Alu"
repeat_region
7051..7172
/rpt_family="Alu"
repeat_region
7192..7500
/rpt_family="Alu"
repeat_region
7518..7820
/rpt_family="Alu"
repeat_region
8207..8500
/rpt_family="Alu"
repeat_region
8576..8878
/rpt_family="Alu"
repeat_region
9023..9154
/rpt_family="Alu"
repeat_region
9165..9470
/rpt_family="Alu"
repeat_region
9887..10188
/rpt_family="Alu"
repeat_region
10221..10342
/rpt_family="L1"
repeat_region
10408..10707
/rpt_family="Alu"
repeat_region
10713..11014
/rpt_family="Alu"
repeat_region
11020..11110
/rpt_family="Alu"
repeat_region
11146..11386
/rpt_family="Alu"
repeat_region
11208..11386
/rpt_family="Alu"
repeat_region
11391..11515
/rpt_family="Alu"
repeat_region
11568..11705
/rpt_family="Alu"

[illegible]

TITLE
 JOURNAL
 REFERENCE
 AUTHORS
 TITLE
 JOURNAL
 REFERENCE
 AUTHORS
 TITLE
 JOURNAL
 COMMENT

Wu C., Wu Y., Wu Y.F., Zhou, J., Zorrilla, S., Nelson, D.,
 Weinstock, G. and Gibbs, R.
 Direct Submission
 2 (bases 1 to 189729)
 Worley, K.C.
 Direct Submission
 Submitted (05-JUN-2000) Human Genome Sequencing Center, Department
 of Molecular and Human Genetics, Baylor College of Medicine, One
 Baylor Plaza, Houston, TX 77030, USA
 3 (bases 1 to 189729)
 Worley, K.C.
 Direct Submission
 Submitted (01-MAR-2002) Human Genome Sequencing Center, Department
 of Molecular and Human Genetics, Baylor College of Medicine, One
 Baylor Plaza, Houston, TX 77030, USA
 On Feb 27, 2002 this sequence version replaced gi:114861669.

----- Genome Center
 Center: Baylor College of Medicine
 Center code: BCM
 Web site: <http://www.hgsc.bcm.tmc.edu/>
 Contact: hgsc-help@bcm.tmc.edu
 ----- Project Information
 Center project name: HBWS
 Center clone name: RP11-144B2
 ----- Summary Statistics
 Sequencing vector: Plasmid; M77789
 Chemistry: Dye-terminator Big Dye; 100% of reads
 Assembly program: Phrap; version 0.990329
 Consensus quality: 191640 bases at least Q40
 Consensus quality: 195135 bases at least Q30
 Consensus quality: 197786 bases at least Q20
 Estimated insert size: 193340; sum-of-contigs estimation
 Quality coverage: 0x in Q20 bases; agarose-fp estimation
 Quality coverage: 11.7x in Q20 bases; sum-of-contigs estimation

 * NOTE: Estimated insert size may differ from sequence length
 (see http://www.hgsc.bcm.tmc.edu/docs/genbank_draft_data.html).
 * NOTE: This is a 'working draft' sequence. It currently
 * consists of 8 contigs. The true order of the pieces
 * is not known and their order in this sequence record is
 * arbitrary. Gaps between the contigs are represented as
 * runs of N, but the exact sizes of the gaps are unknown.
 * This record will be updated with the finished sequence
 * as soon as it is available and the accession number will
 * be preserved.
 * 1 64773: contig of 64773 bp in length
 * 64774 64873: gap of unknown length
 * 64874 99097: contig of 34224 bp in length
 * 99098 99197: gap of unknown length
 * 99198 135331: contig of 36134 bp in length
 * 135332 135431: gap of unknown length
 * 135432 161344: contig of 25913 bp in length
 * 161345 161444: gap of unknown length
 * 161445 171231: contig of 9787 bp in length
 * 171232 171331: gap of unknown length
 * 171332 180922: contig of 9591 bp in length
 * 180923 181022: gap of unknown length
 * 181023 187271: contig of 6249 bp in length
 * 187272 187371: gap of unknown length
 * 187372 189729: contig of 2358 bp in length.
 Location/Qualifiers
 1. 189729
 /organism="Homo sapiens"
 /mol_type="genomic DNA"
 /db_xref="taxon:9606"
 /chromosome="12q"
 /clone="RP11-144B2"
 64774..64873
 /estimated_length=unknown
 99098..99197
 /estimated_length=unknown

gap 135332..135431
 /estimated_length=unknown
 gap 161345..161444
 /estimated_length=unknown
 gap 171232..171331
 /estimated_length=unknown
 gap 180923..181022
 /estimated_length=unknown
 gap 187272..187371
 /estimated_length=unknown
 ORIGIN
 Query Match 99.6%; Score 897.8; DB 14; Length 189729;
 Best Local Similarity 99.8%; Pred. No. 2.8e-171;
 Matches 899; Conservative 0; Mismatches 2; Indels 0; Gaps 0;
 QY 1 TGTTCATATTAATTAATAAATCACTCAGTTGGGCACTCAAGCCTGTAACACAGT 60
 Db 172643 TGTTCATATTAATTAATAAATCACTCAGTTGGGCACTCAAGCCTGTAACACAGT 172702
 QY 61 ACTTTGGAAGTCCCAAGGTGGGTGGATCCTTGGAGTGAAGATTTCGAGACCAGCTGGTC 120
 Db 172703 ACTTTGGAAGTCCCAAGGTGGGTGGATCCTTGGAGTGAAGATTTCGAGACCAGCTGGTC 172762
 QY 121 AATATGGTGAACCCCTATCTCTACTAAAAATACAAAAATTAGCTGGGTGTAGTGCAT 180
 Db 172763 AATATGGTGAACCCCTATCTCTACTAAAAATACAAAAATTAGCTGGGTGTAGTGCAT 172822
 QY 181 GCCTGTAGTCCAGCTACTCGGGAGGCTGAGGCAAGAGAATTCCTTGAACCTGGGAGCA 240
 Db 172823 GCCTGTAGTCCAGCTACTCGGGAGGCTGAGGCAAGAGAATTCCTTGAACCTGGGAGCA 172882
 QY 241 GAGGTTCAGTGAGCCGAGATCCCACTCCAGCTCCAGCTGGGCGACACAGCCGAGACT 300
 Db 172883 GAGGTTCAGTGAGCCGAGATCCCACTCCAGCTCCAGCTGGGCGACACAGCCGAGACT 172942
 QY 301 CTATCTCAAAAAATAAATAAATAAAGGATCGGAGAGAAACAAAACTAATAAGAT 360
 Db 172943 CTATCTCAAAAAATAAATAAATAAAGGATCGGAGAGAAACAAAACTAATAAGAT 173002
 QY 361 TCCTGAAGGTGAAGCAGATAGTAATATATGTAATAAAGTTTAAATGCAATTTAACT 420
 Db 173003 TCCTGAAGGTGAAGCAGATAGTAATATATGTAATAAAGTTTAAATGCAATTTAACT 173062
 QY 421 GTAATCTTATTTGTTTATTTGTTTATATAAAGTAAACCAAGCAAAAGTAAATGCAATTTCAA 480
 Db 173063 GTAATCTTATTTGTTTATTTGTTTATATAAAGTAAACCAAGCAAAAGTAAATGCAATTTCAA 173122
 QY 481 ACTCTACATAAATATCTATTATGGAAGTGAAGGCACTTATAATCTCTACTACCCAAAGA 540
 Db 173123 ACTCTACATAAATATCTATTATGGAAGTGAAGGCACTTATAATCTCTACTACCCAAAGA 173182
 QY 541 TAACCAAGTTACATAATTCCTCCAGATTTTGGGGCATACTAGCTTTTATTTTGGGAA 600
 Db 173183 TAACCAAGTTACATAATTCCTCCAGATTTTGGGGCATACTAGCTTTTATTTTGGGAA 173242
 QY 601 AATTTCCATGTCAGGCATACCTAATTTTCTAATATGCTAGTATTTCCATTTAAGG 660
 Db 173243 AATTTCCATGTCAGGCATACCTAATTTTCTAATATGCTAGTATTTCCATTTAAGG 173302
 QY 661 ATGTTTCCATAATTTTAAAAATACATGCTTTAAAGTAGAGAAACTAGGTTGGGCATGGTG 720
 Db 173303 ATGTTTCCATAATTTTAAAAATACATGCTTTAAAGTAGAGAAACTAGGTTGGGCATGGTG 173362
 QY 721 CTCAGCCTGTATCCAGCACTTTGGGAGCGGAGCAAAATGATCCTTTGAGGTCGGCA 780
 Db 173363 CTCAGCCTGTATCCAGCACTTTGGGAGCGGAGCAAAATGATCCTTTGAGGTCGGCA 173422
 QY 781 GTTTGAGACCGCCTGGCAACATGATGAAACACCCCTCTCTAATAAAAAATACAAAAATTA 840
 Db 173423 GTTTGAGACCGCCTGGCAACATGATGAAACACCCCTCTCTAATAAAAAATACAAAAATTA 173482
 QY 841 GCTGGGCATGGTGGCAACACCTGTAGTCCCAAGTCTCAGGAGTCTGAGGCGAGGAGTAT 900

FEATURES
 source
 gap
 gap

```
Db 173483 GCTGGGCATGGTGGCGAGCACCTGTAGTCCCGACTAGGAGTCTGAGGCGAGGATAT 173542
|||||
QY 901 C 901
Db 173543 C 173543

RESULT 6
AL390961
LOCUS DEFINITION
AL390961 173950 bp DNA linear PRI 18-MAY-2005
Human DNA sequence from clone Rp13-16H11 on chromosome 10. Contains
three novel genes, the gene for trans-prenyltransferase (TPT), the
3' end of the SSH3BP gene for spectrin SH3 domain binding protein
1, a heat shock protein pseudogene and a CpG island, complete
sequence.
ACCESSION
AL390961
VERSION AL390961.17 GI:19351892
KEYWORDS HTG; CpG island; SSH3BP; TPT.
SOURCE Homo sapiens
ORGANISM
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini;
Hominae; Homo.
1 (bases 1 to 173950)
Direct Submission
Submitted (13-MAY-2005) Wellcome Trust Sanger Institute, Hinxton,
Cambridgeshire, CB10 1SA, UK. E-mail enquiries: vegas@sanger.ac.uk
Clone requests: clonerequest@sanger.ac.uk
On Mar 11, 2002 this sequence version replaced gi:18121480.
The following abbreviations are used to associate primary accession
numbers given in the feature table with their source databases:
Em: EMBL; Sw: SWISSPROT; Tr: TrEMBL; Wp: WORMPEP; Information
on the WORMPEP database can be found at
http://www.sanger.ac.uk/Projects/C_elegans/wormpep
This sequence
was generated from part of bacterial clone contigs of human
chromosome 10, constructed by the Sanger Centre Chromosome 10
Mapping Group. Further information can be found at
http://www.sanger.ac.uk/HGP/Chr10
Rp13-16H11 is from the library RPCI-13.1 constructed by the group
of Pieter de Jong. For further details see
http://www.chori.org/bacpac/home.htm
VECTOR: pBACe3.6
----- Genome Center
Center: Wellcome Trust Sanger Institute
Center code: SC
Web site: http://www.sanger.ac.uk
Contact: vegas@sanger.ac.uk
-----
This sequence was finished as follows unless otherwise noted: all
regions were either double-stranded or sequenced with an alternate
chemistry or covered by high quality data (i.e., phred quality >=
30); an attempt was made to resolve all sequencing problems, such
as compressions and repeats; all regions were covered by at least
one subclone; and the assembly was confirmed by restriction digest,
except on the rare occasion of the clone being a YAC.
FEATURES
Location/Qualifiers
1..173950
/organism="Homo sapiens"
/mol_type="genomic DNA"
/db_xref="taxon:9606"
/chromosome="10"
/clone="RP13-16H11"
/clone_lib="RPCI-13.1"
2000
misc_feature
/note="Clone_right_end: RP11-128B16"
join(32186..32554,33872..33974)
gene
/locus_tag="RP13-16H11.7-001"
join(32186..32554,33872..33974)
mRNA
/locus_tag="RP13-16H11.7-001"
/product="putative novel transcript"
/note="match: ESTs: AA888857.1"

join(36944..37240,37909..38008,38349..38551,39512..39749,
42023..42404)
/locus_tag="RP13-16H11.1-001"
mRNA
join(36944..37240,37909..38008,38349..38551,39512..39749,
42023..42404)
/locus_tag="RP13-16H11.1-001"
/product="novel transcript"
/note="match: ESTs: AA429709.1 AL042653.1 AL043087.2"
join(37058..37240,38349..38551,39512..39615,42023..42412)
/locus_tag="RP13-16H11.1-002"
gene
join(37058..37240,38349..38551,39512..39615,42023..42412)
/locus_tag="RP13-16H11.1-002"
/product="novel transcript"
/note="match: ESTs: AI183567.1 AL040562.1"
join(38290..38551,39512..39615,42023..42069)
/locus_tag="RP13-16H11.1-003"
mRNA
join(38290..38551,39512..39615,42023..42069)
/locus_tag="RP13-16H11.1-003"
/product="novel transcript"
/note="match: ESTs: AA412640.1"
42384..42389
polyA_signal
/locus_tag="RP13-16H11.1-002"
42407
polyA_site
/locus_tag="RP13-16H11.1-002"
complement(join(53164..53350,57365..57452,65074..65117))
gene
/locus_tag="RP13-16H11.2-001"
complement(join(53164..53350,57365..57452,65074..65117))
mRNA
/locus_tag="RP13-16H11.2-001"
/product="novel transcript"
/note="match: ESTs: BG460048.1 BG460049.1"
join(86617..86798,91120..91152,93635..93699,94244..94352,
98596..98726,109176..109317,112764..112875,112972..113081,
124198..124278,124424..124537,131455..131535,
135291..135756)
gene
/locus_tag="RP13-16H11.3"
/locus_tag="RP13-16H11.3-001"
mRNA
join(86617..86798,91120..91152,93635..93699,94244..94352,
98596..98726,109176..109317,112764..112875,112972..113081,
124198..124278,124424..124537,131455..131535,
135291..135756)
/locus_tag="RP13-16H11.3"
/product="trans-prenyltransferase (TPT)"
/note="match: ESTs: AA889371.1 AL558274.1 BF982221.1
BI255737.1 BI761095.1 BM824056.1 BM994354.1 BQ018783.1
BQ212558.1
match: CDNA: AF118395.1 AF118855.1 AK012307.1 AK024802.1
BC026820.1"
join(86670..86798,91120..91152,93635..93699,94244..94352,
98596..98726,109176..109317,112764..112875,112972..113081,
124198..124278,124424..124537,131455..131535,
135291..135431)
CDS
/locus_tag="RP13-16H11.3"
/locus_tag="RP13-16H11.3-001"
/standard_name="OTTHUMP00000019346"
/notes="match: proteins: Q9CZQ1 Q9WU69 Q9Y2W5"
/codon_start=1
/product="trans-prenyltransferase (TPT)"
/protein_id="CAI17280.1"
/db_xref="GI:55959523"
/db_xref="GOA:Q5T2R2"
/db_xref="InterPro:IPR000092"
/translation="MASRWRRRGSGWKAARSPGSGPRAGPLGPSAAAEVRAQV
HRRKGLDLSQIPYINLVKHLTSACPNVIRISRFHTTDFDTSKTHSGEYKTFPKLGMWD
LKGLYEDIRKELLISTSELKSEYFYFDGKGAFRIIIVAMARACNHNHNRHVOA
SQRAIALIAEMIHTASLIVDDVDIDASSRRGKHTVNVKGEKAVLAGDILISAASIA
LARIGNTVTISLTVQIEDLVRCGLQSGKENENEFAYLEKTFKKTASLIANSCK
AVSLGCPDVPVHEIAVQYGVKNVGFIAQLIDVDLDTSCSDQMGKTSADLKLGLATG
PVLFAQQCFPMNAMIWRPFLSGPDVDRARQYVLQSDGVQQTITLAAQYCHEAIREIS
KLRPSPERDALIQLSEIVLTRDK"
join(94266..94352,98596..98726,109093..109317,
112764..112875,112972..112977)
```



```
/gene="RP13-16H11.3"
/locus_tag="RP13-16H11.3-004"
join(94266..94352,98596..98726,109093..109317,
112764..112875,112972..112977)
/gene="RP13-16H11.3"
/locus_tag="RP13-16H11.3-004"
/product="trans-prenyltransferase (TPT)"
/notes="match: ESTs: AA362710.1 A1873444.1 B1062674.1"
join(103580..103675,103811..104363,104660..104790)
/locus_tag="RP13-16H11.6-001"
/pseudo
join(103580..103675,103811..104363,104660..104790)
/locus_tag="RP13-16H11.6-001"
/notes="match: proteins: AAC84170 AH35665 O62564 O86866
O93240 P02827 P16627 P17066 P17879 P29843 P55063 P79984
Q05944 Q28222 Q8UW6 Q8WQ17 Q91233 Q925V6 Q9NGK9 Q9OWU5
Q9TUC3 Q9XZ32"
/pseudo
/codon_start=1
/product="heat shock protein pseudogene"
join(106593..107031,107691..107736)
/locus_tag="RP13-16H11.5-001"
join(106593..107031,107691..107736)
/locus_tag="RP13-16H11.5-001"
/product="putative novel transcript"
/notes="match: ESTs: BF387222.1"
join(<109301..109317,112764..112875,112972..113081,
124198..124278,124424..124537,129603..129775,
131455..131535,135291..135499)
/gene="RP13-16H11.3"
/locus_tag="RP13-16H11.3-002"
join(<109301..109317,112764..112875,112972..113081,
124198..124278,124424..124537,129603..129775,
131455..131535,135291..135499)
/gene="RP13-16H11.3"
/locus_tag="RP13-16H11.3-002"
/product="trans-prenyltransferase (TPT)"
/notes="match: ESTs: BM806663.1"
join(<109301..109317,112764..112875,112972..113081,
124198..124278,124424..124537,129603..129644)
/gene="RP13-16H11.3"
/locus_tag="RP13-16H11.3-002"
/standard_name="O1THUMP00000019347"
/codon_start=3
/product="trans-prenyltransferase (TPT)"
/protein_id="CA117282.1"
/db_xref="GI:55959524"
/db_xref="GOA:Q5T2R1"
/db_xref="InterPro:IPR000092"
/db_xref="UniProt/TREMBL:Q5T2R1"
/translation="MGKKAVLAGDILISAASIALARIQNTTIVISILTOVIEDLVKRG
FIQLGSKENENRFAHYLEKTFKTLASLANSKAVSVILGCPDPVVEIAYQYGNVG
IAFLQDDVLDFTSCSDQMGKPTSDALKGLATPVLFCQVHCQCGTKGCKKH"
join(123609..123681,124198..124278,124424..124537,
131455..131535,135291..135740)
/gene="RP13-16H11.3"
/locus_tag="RP13-16H11.3-003"
join(123609..123681,124198..124278,124424..124537,
131455..131535,135291..135740)
/gene="RP13-16H11.3"
/locus_tag="RP13-16H11.3-003"
/product="trans-prenyltransferase (TPT)"
/notes="match: ESTs: B1756998.1"
join(complement(AL139404.9:77755..77994),
37.48; Score 337.2; DB 8; Length 173950;
Query Match 63.8; Pred. No. 2.8e-58;
Best Local Similarity 0; Mismatches 313; Indels 4; Gaps 3;
Matches 559; Conservative 0;

30 GGGCAGTGCACCTGTAACACAGTCTTTTGGAAAGTCCAGGTGGGTGGATCAC 89
|||||
133844 GGGCGCAGTGGCTCACACCTGTATCGCAGCATTTTGGGAGGCCAAGGTGGGTGGATCAC 133903
|||||
```

```
QY 90 TTGAGGTGAGAAGTTTCGAGACCAGCCCTGGTCAATATGGTGAACCCCTATCTCTACTAAAA 149
|||||
Db 133904 TTGAGGTGAGAGTTTTCGAGACCAGCCCTGGCCAAACATGGTGAACCCCTCTCTACTAAAA 133963
|||||
QY 150 ATACAAAAATTTAGCTGGGTGTAGTGATGCATGCCCTGTAGTCCCGAGCTACTCGGAGGCTG 209
|||||
Db 133964 ATACAAACATTTAGCCAGGCGTGGTAGTGCACGCCCTATAATCTTAGCTACTTTAGGAGGCTG 134023
|||||
QY 210 AGGCAAGAGAAATTCCTTGAACCTGGGAGGAGAGGTTGCGAGTGCAGCCGAGATCCACCAC 269
|||||
Db 134024 AGGCACAGAAATCCCTTTGAGCCTGGGAGGAGAGGTTGCGAGTGCAGCCAGATCACGCCAC 134083
|||||
QY 270 TGCACTCCAGCTCGGGCGACACAGCGAGACTCTATCTCAAAAAATAAATAAATAAATA 329
|||||
Db 134084 TGCAATTCAGCTCGGGCGACAGAGCAAGATTCATCTCAAAAAATAAATAAATAAATAA 134143
|||||
QY 330 AAGGATCGGAGAGAAAAAACAACCTAATAAGATTCCTGAAGGTGACGAGAG-ATACGTAAT 388
|||||
Db 134144 ATAAATGTCAGAGGGATCATGAGCCATCGAGCAAGGAAATAGAAATTATAGTATATGACTTG 134203
|||||
QY 389 TATATGTAATAAAGTTTAAATGCAATTTTAACGTGAATCTTATGTTTATTTTGGTTATAA 448
|||||
Db 134204 ATTCGCAATAAATATTTTACATCATAGAAATGATACCTTTATAGTATTTGGCTTTTAG 134263
|||||
QY 449 AAGTAAACAGCCAAAAGTAAATGCAACTTCAAACTCTACATAAAATATCTATATGGAAG 508
|||||
Db 134264 AAAGAACCAATATTTTAAACATGGTTATAGAAATAGGATGTAATATCTATAGTCTAACA 134323
|||||
QY 509 TGGAGGCACTCTAATATCTACTACCCAAAGATTAACAGATTACATATATCTCCAGATTTT 568
|||||
Db 134324 AACAGTATAGTTTAAAGGAAAGCAAGAGCTCATAGAGATAAAATCTGAAGTTGATGCT 134383
|||||
QY 569 TGGGGCATACTAGCTTTTATTTTGGGAAATTTTCATGTCAGGCATACCTAAATTT 628
|||||
Db 134384 ACCTAACAGAAATTAACAACAAAACAATAAGACAGCAGTGTGATGATGAGCGGCACA 134443
|||||
QY 629 TTCTAAATGTCATGTAGTATTTCCATTTAAGAGATGTTCCATAAATTTTAAATATACATGCT 688
|||||
Db 134444 TTGTTGGTGGCTAAACCCCTCATCTTAACAAAAGGTAACAGGTAATAAATAAATTTGCT 134503
|||||
QY 689 --TTAAAGTAGAGAAACTAGTGTGGCATGTTGGTGTCAAGCTGT-ATCCGAGCACTTTG 745
|||||
Db 134504 AATCAAGAAATAAATAAATAGGCTGGCCAGCTGCTCACACCTGTAATCCCGACACTTTG 134563
|||||
QY 746 GGAGCCGAGCAAAATGATCACTTTGAGTCCGGAGTTTGAACAGCAGCTGGACACATG 805
|||||
Db 134564 GGAGCCCAAGGCAGGCAGATCACTTTGAGGCCAGGAGTTGCGAGCAGCTGGCCCAATG 134623
|||||
QY 806 ATGAACACCCCTCTCTAATAAATAACAAAAATTAGCTGGGCATGTTGGCAAGCACCTGT 865
|||||
Db 134624 ACAAAACCCCATCTCTACTAATAAATAACAAAAATTAGCCGGGCATGTTGGCAGCGCTGT 134683
|||||
QY 866 AGTCCAGCTACTCAGGAGTCTGAGCGCAGGAGTATC 901
|||||
Db 134684 AGTCCAGCTACTTGGGAGGCTGAGGCAGGACAATC 134719
|||||

RESULT 7
AC022245/c
LOCUS
DEFINITION
Homo sapiens chromosome 10 clone RP11-356G5 map 10, WORKING DRAFT
SEQUENCE, 13 unordered pieces.
AC022245
VERSION
HTG; HTGS PHASE1; HTGS DRAFT.
KEYWORDS
SOURCE
Homo sapiens (human)
ORGANISM
Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini;
Homnidae; Homo.
REFERENCE
1 (bases 1 to 174937)
AUTHORS
Birren,B., Linton,L., Nusbaum,C. and Lander,E.
TITLE
Homo sapiens chromosome 10, clone RP11-356G5
```

JOURNAL
REFERENCE
AUTHORS

Unpublished
2 (bases 1 to 174937)
Birren,B., Linton,L., Nusbaum,C., Lander,E., Abraham,H., Allen,N.,
Anderson,S., Baldwin,J., Barna,N., Beckerly,R., Bada,F.,
Boguslavskiy,L., Boukhgalter,B., Brown,A., Burkett,G., Castle,A.,
Choepl,Y., Colangelo,M., Collins,S., Collymore,A., Cooke,P.,
DeArellano,K., Dewar,K., Domino,M., Doyle,M., Fencstator,J.,
Perreira,P., Fitzhugh,W., Forrest,C., Gage,D., Galagan,J.,
Gardyna,S., Grant,G., Hagos,B., Heatford,A., Horton,L.,
Howland,J.C., Johnson,R., Jones,C., Kann,L., Karatas,A., Klein,J.,
Landers,T., Lehoczyk,J., Levine,R., Liu,C., Liu,G., Locke,K.,
MacDonald,P., Marquis,N., McEwan,P., McGurk,A., McKernan,K.,
McPheeters,R., Meldrim,J., Meneus,L., Morrow,J., Naylor,J.,
Norman,C.H., O'Connor,T., O'Donnell,P., Olivart,T.M., Peterson,K.,
Pierre,N., Pisani,C., Pollara,V., Raymond,C., Riley,R., Rothman,D.,
Roy,A., Santos,R., Severy,P., Spencer,B., Stange-Thomann,N.,
Stojanovic,N., Subramanian,A., Talamas,J., Tesfaye,S., Theodore,J.,
Tirrell,A., Vassiliev,H., Viel,R., Vo,A., Wu,X., Wyman,D., Ye,W.J.,
Zimmer,A. and Zody,M.
Direct Submission
Submitted (27-JAN-2000) Whitehead Institute/MIT Center for Genome
Research, 320 Charles Street, Cambridge, MA 02141, USA
3 (bases 1 to 174937)
Birren,B., Linton,L., Nusbaum,C., Lander,E., Allen,N., Anderson,S.,
Barna,N., Bastien,V., Boguslavskiy,L., Boukhgalter,B., Brown,A.,
Camarata,J., Campopiano,A., Chang,J., Choepl,Y., Colangelo,M.,
Collins,S., Collymore,A., Cooke,P., DeArellano,K., Dewar,K.,
Diaz,J.S., Dodge,S., Faro,S., Ferreira,P., Fitzhugh,W., Gage,D.,
Galagan,J., Gardyna,S., Ginde,S., Goyette,M., Graham,L.,
Grand-Pierre,N., Hagos,B., Heatford,A., Horton,L., Hulme,W.,
Iliev,I., Johnson,R., Jones,C., Karatas,A., Horton,L., Liu,G.,
Lamazares,R., Landers,T., Lehoczyk,J., Levine,R., Liu,G.,
MacLean,C., Macdonald,P., Marquis,N., Matthews,C., McCarthy,M.,
McEwan,P., McKernan,K., McPheeters,R., Meldrim,J., Meneus,L.,
Mihova,T., Mlenca,V., Murphy,T., Naylor,J., Nguyen,C., Norbu,C.,
Norman,C.H., O'Connor,T., O'Donnell,P., O'Neill,D., Oliver,J.,
Peterson,K., Phunkhang,P., Pierre,N., Pollara,V., Raymond,C.,
Retta,R., Rieback,M., Riley,R., Rise,C., Rogov,P., Roman,J.,
Rosetti,M., Roy,A., Santos,R., Schauer,S., Schuback,R., Seaman,S.,
Severy,P., Sougnez,C., Spencer,B., Stange-Thomann,N.,
Stojanovic,N., Strauss,N., Subramanian,A., Talamas,J., Tesfaye,S.,
Theodore,J., Travers,M., Travis,N., Trigilio,J., Vassiliev,H.,
Viel,R., Vo,A., Wilson,B., Wu,X., Wyman,D., Ye,W.J., Young,G.,
Zainoun,J., Zembek,L., Zimmer,A. and Zody,M.
Direct Submission
Submitted (24-AUG-2002) Whitehead Institute/MIT Center for Genome
Research, 320 Charles Street, Cambridge, MA 02141, USA
On Mar 12, 2000 this sequence version replaced gi:6850466.
All repeats were identified using RepeatMasker:
Smit, A.F.A. & Green, P. (1996-1997)
<http://ftp.genome.washington.edu/RM/RepeatMasker.html>

TITLE
JOURNAL
COMMENT

Center: Whitehead Institute/ MIT Center for Genome Research
Center code: WIBR
Web site: <http://www-seq.wi.mit.edu>
Contact: sequence_submissions@genome.wi.mit.edu
----- Project Information
Center project name: L4212
Center clone name: 356_G_5
----- Summary Statistics
Sequencing vector: M13; M77815; 100% of reads
Chemistry: Dye-terminator Big Dye; 100% of reads
Assembly program: Phrap; version 0.960731
Consensus quality: 167019 bases at least Q40
Consensus quality: 170472 bases at least Q30
Consensus quality: 171999 bases at least Q20
Insert size: 176000; agarose-fp
Insert size: 173737; sum-of-contigs
Quality coverage: 5.0 in Q20 bases; agarose-fp
Quality coverage: 5.1 in Q20 bases; sum-of-contigs

* NOTE: This is a 'working draft' sequence. It currently
* consists of 13 contigs. The true order of the pieces

* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.

	1	1785:	contig of 1785 bp in length
		1786:	gap of 100 bp
		1885:	contig of 1226 bp in length
		3111:	gap of 100 bp
		3211:	contig of 2411 bp in length
		5622:	gap of 100 bp
		5722:	contig of 4869 bp in length
		10591:	gap of 100 bp
		10691:	contig of 4722 bp in length
		15413:	gap of 100 bp
		15513:	contig of 9534 bp in length
		25047:	gap of 100 bp
		25147:	contig of 11423 bp in length
		36570:	gap of 100 bp
		36671:	contig of 13593 bp in length
		50264:	gap of 100 bp
		50364:	contig of 13919 bp in length
		64282:	gap of 100 bp
		64382:	contig of 20390 bp in length
		84772:	gap of 100 bp
		84873:	contig of 20350 bp in length
		105223:	gap of 100 bp
		105322:	contig of 32200 bp in length
		137522:	gap of 100 bp
		137523:	contig of 37315 bp in length.
FEATURES	Location/Qualifiers		
source	1..174937		
	/organism="Homo sapiens"		
	/mol_type="genomic DNA"		
	/db_xref="taxon:9606"		
	/chromosome="10"		
	/map="10"		
	/clone="RP11-356G5"		
	/clone_lib="RPC1-11 Human Male BAC"		
misc_feature	1..1785		
	/notes="assembly_fragment"		
gap	1786..1885		
	/estimated_length=100		
misc_feature	1886..3111		
	/note="assembly_fragment"		
gap	3112..3211		
	/estimated_length=100		
misc_feature	3212..5622		
	/notes="assembly_fragment"		
gap	5623..5722		
	/estimated_length=100		
misc_feature	5723..10591		
	/notes="assembly_fragment"		
gap	10592..10691		
	/estimated_length=100		
misc_feature	10692..15413		
	/note="assembly_fragment"		
gap	15414..15513		
	/estimated_length=100		
misc_feature	15514..25047		
	/note="assembly_fragment"		
gap	25048..25147		
	/estimated_length=100		
misc_feature	25148..36570		
	/notes="assembly_fragment"		
gap	36571..36670		
	/estimated_length=100		
misc_feature	36671..50263		
	/notes="assembly_fragment"		
gap	50264..50363		
	/estimated_length=100		
misc_feature	50364..64282		

/note="assembly_fragment"
64283. .64382
/estimated_length=100
64383. .84772
/note="assembly_fragment"
84773. .84872
/estimated_length=100
84873. .105222
/note="assembly_fragment"
105223. .105322
/estimated_length=100
105323. .137522
/note="assembly_fragment"
clone_end:T7
vector_side:right
137523. .137622
/estimated_length=100
137623. .174937
/note="assembly_fragment"
clone_end:SP6
vector_side:left"

ORIGIN

Query Match 37.4%; Score 337.2; DB 14; Length 174937;
Best Local Similarity 63.8%; Pred. No. 2.8e-58;
Matches 559; Conservative 0; Mismatches 313; Indels 4; Gaps 3;

QY 30 GGGCACAGTACTCAAGCTGTAAACACAGACTTTTGGAACTCCAAAGTGGGTGGATCAC 89
DB 50262 GGGCCAGTGGCTCACACTGTATCGCAGCATTTTGGAGGCCAAGTGGGTGGATCAC 50203

QY 90 TTGAGTGGAGAAATTCGAGACAGCCTGGTCAATATGGTGAACCTATCTCTACTAAAA 149
DB 50202 TTGAGTTCAGGATTTTGAGCAGCCTGGCCAACTGTGAAACCCCTCTCTACTAAAA 50143

QY 150 ATACAAAAATAGCTGGGTGTAGTATGATGATGCTGTAGTCCAGCTACTCGGAGGCTG 209
DB 50142 ATACAAACATTAGCCAGGCGTGTAGTGCACGCTTAAATCTAGTACTTAGGAGGCTG 50083

QY 210 AGGCAAGAAATTTGCTTGAACCTGGAGGCAGAGGTTCAGTGGAGCCAGATCCCAACCAC 269
DB 50082 AGGCAGAGAAATCCCTTGAGCTGGAGGCAGAGGTTCAGTGGAGCCAGATCCCAACCAC 50023

QY 270 TGCACTCCAGCTGGCGCACAGCGAGACTCTATCTCAAAAAATAAATAAATAAATA 329
DB 50022 TGCACTCCAGCTGGCGCACAGCGAGACTCTATCTCAAAAAATAAATAAATAAATA 49963

QY 330 AAGGATCGAGAGAAACAAACTATAGATTCCTGAAGGTAAAGCAGAG-ATACGTAAT 388
DB 49962 ATAAATGTCAGAGGGATCATGAGCCATGGAGGAAGAAATAGAAATTATAGTATGACTTG 49903

QY 389 TATATGTAATAAGTTTAAATGCAATTTTAACTGTAATCTTATTGTTTATTTTGGTTATAA 448
DB 49902 ATTCTCAATAAATATTTTACATCATAGAAATGATCTTTATAGTATTTGCTTTTAG 49843

QY 449 AAGTAACCAAGCCAAAGTAATGCAACTTCAAACTCTACATAAATATCTATTATGGAAG 508
DB 49842 AAGAACCAATATTTTAAACATGGTTATAGAATAGGATGTAATATCTATAGTCTAACA 49783

QY 509 TGAAGGCATCTATAATCTCTATACCAAGATAACCGATTACATATTCCTCCAGATTTT 568
DB 49782 AACAGTATAGTTTAAAGGAAGGACAGAGCTCATAGAAGATAAATCTGAAGTTGATGCT 49723

QY 569 TGGGCATACACTAGCTTTTATTTTGGGAATTTCCATGTGAGGCATACCTAATTT 628
DB 49722 ACTTAACAGAAATGTAATCAACAAACAACTTAAGACAGAGTGTGATGATGAGCGGCACA 49663

QY 629 TTCTAAATGCTATGTAGTATTTCCATTTTAAGGATGTTTCCATATATTTTAAAAATACATGCT 688
DB 49662 TTCTGGGTGGCTAAACCCCTCATCTTAAACAAAGGTAACAGGTAATAACTAAATTTGTT 49603

QY 689 --TTAAAGTAGAAACTAGGTTGGGATGGTGGCTCAGCCCTGT-ATCCAGCACTTTG 745

Db 49602 AATCAAGAAATAAAAACTAGGCTGGCAGCGTGGCTCACACCTGTATATCCACACCTTTG 49543
QY 746 GAGGCGGAGGCAATGGATCACTTCAAGTCCGAGTTTGAGACCAAGCTTGGCAACATG 805
Db 49542 GAGGCGCAAGGCGAGGAGATCACTTGAGGCGAGAGTTCGAGACCAAGCTTGGCAACATG 49483
QY 806 ATCAAAACACCCCTCTCTAATAAAAAATACAAAAATAGCTGGGCATGTGGCAAGCACCTGT 865
Db 49482 ACAAACCCCATCTCTACTTAAATAATACAAAAATAGCCGGGCATGTGGCAGCGCCTGT 49423
QY 866 AGTCCAGCTACTTGGAGGCTGAGGAGTCTGAGCAGGAGTATC 901
Db 49422 AGTCCAGCTACTTGGAGGCTGAGGAGTCTGAGCAGGAGTATC 49387

RESULT 8

AL591164/c

LOCUS

DEFINITION

14 unordered pieces.

ACCSSION

AL591164.3

VERSION

HTG; HTGS PHASE1; HTGS DRAFT.

KEYWORDS

SOURCE

ORGANISM

Homo sapiens

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini;

Hominidae; Homo.

1

REFERENCE

Burton, J.

Direct Submission

Submitted (18-DEC-2001) Wellcome Trust Sanger Institute, Hinxton,

Cambridgeshire, CB10 1SA, UK. E-mail enquiries:

humquery@sanger.ac.uk Clone requests: clonerequest@sanger.ac.uk

On Dec 20, 2001 this sequence version replaced gi:14575431.

----- Genome Center

Center: Wellcome Trust Sanger Institute

Center code: SC

Web site: http://www.sanger.ac.uk

Contact: humquery@sanger.ac.uk

----- Project Information

Center project name: BB31H8

----- Summary Statistics

Assembly program: XGAP4; version 4.5

Sequencing vector: plasmid; 108752; 100% of reads

Chemistry: Dye-terminator Big Dye; 100% of reads

Consensus quality: 190732 bases at least Q40

Consensus quality: 193432 bases at least Q30

Consensus quality: 194855 bases at least Q20

Insert size: 195945; sum-of-contigs

Insert size: 204159; 4.1% error; agarose-fp

Quality coverage: 4.66x in Q20 bases; sum-of-contigs Quality

coverage: 4.47x in Q20 bases; agarose-fp

* NOTE: This is a 'working draft' sequence. It currently

* consists of 14 contigs. The true order of the pieces

* is not known and their order in this sequence record is

* arbitrary. Gaps between the contigs are represented as

* runs of N, but the exact sizes of the gaps are unknown.

* This record will be updated with the finished sequence

* as soon as it is available and the accession number will

* be preserved.

* 1

* 6607: contig of 6607 bp in length

* 6708: gap of 100 bp

* 13612: contig of 6905 bp in length

* 13613: gap of 100 bp

* 13713: contig of 25583 bp in length

* 39295: gap of 100 bp

* 39296: contig of 13252 bp in length

* 52647: gap of 100 bp

* 52748: contig of 2315 bp in length

* 55063: gap of 100 bp

* 73956: contig of 18794 bp in length

* 55163

* 73957 74056: gap of 100 bp
 * 74057 94138: contig of 20082 bp in length
 * 94139 94238: gap of 100 bp
 * 118037: contig of 23799 bp in length
 * 118038 118137: gap of 100 bp
 * 118138 120149: contig of 2012 bp in length
 * 120150 120449: gap of 100 bp
 * 120450 174347: contig of 54098 bp in length
 * 174348 174447: gap of 100 bp
 * 174448 177356: contig of 2909 bp in length
 * 177357 177456: gap of 100 bp
 * 177457 180727: contig of 3271 bp in length
 * 180728 180827: gap of 100 bp
 * 180828 194321: contig of 13494 bp in length
 * 194322 194421: gap of 100 bp
 * 194422 197245: contig of 2824 bp in length.

FEATURES

source

Location/Qualifiers
 1..197245
 /organism="Homo sapiens"
 /mol_type="genomic DNA"
 /db_xref="taxon:9606"
 /chromosome="10"
 /clone="RP13-11H8"
 /clone_lib="RPC1-13.1"

misc_feature
 1..6607
 /note="assembly_fragment:00616
 fragment_chain:1"
 6708..13612
 /note="assembly_fragment:01749
 fragment_chain:1"
 13713..39295
 /note="assembly_fragment:00559
 fragment_chain:1"
 39396..52647
 /note="assembly_fragment:01954
 fragment_chain:1"
 52748..55062
 /note="assembly_fragment:00859
 fragment_chain:1"
 55163..73956
 /note="assembly_fragment:01247
 fragment_chain:1"
 74057..94138
 /note="assembly_fragment:01887
 fragment_chain:1"
 94239..118037
 /note="assembly_fragment:00055
 fragment_chain:1"
 118138..120149
 /note="assembly_fragment:00714
 fragment_chain:2"
 120250..174347
 /note="assembly_fragment:00818
 fragment_chain:2"
 174448..177356
 /note="assembly_fragment:00335
 fragment_chain:2"
 177457..180727
 /note="assembly_fragment:01862"
 180828..194321
 /note="assembly_fragment:01417
 fragment_chain:3"
 194422..197245
 /note="assembly_fragment:01084
 fragment_chain:3
 clone_end:T7
 vector_side:right"

ORIGIN

Query Match 37.4%; Score 337.2; DB 14; Length 197245;
 Best Local Similarity 63.8%; Pred. No. 2.7e-58;
 Matches 559; Conservative 0; Mismatches 313; Indels 4; Gaps 3;

Qy 30 GGGCAGTGTGCTCAAGCTGTAAACACAGTACTTTGGAAAGTCCAAGTGGGTGGATCAC 89
 Db 115022 GGGCGAGTGGCTCACACCTGTAAATCGCAGATTTTGGAGCCCAAGTGGTGGATCAC 114963
 Qy 90 TTGAGGTGAGAGTTTCGAGACCGCTTGGTCAATATATGTTGAACCCCTATCTCTACTAAAA 149
 Db 114962 TTGAGGTGAGAGTTTCGAGACCGCTTGGCCCAACATGTTGAACCCCGTCTCTACTAAAA 114903
 Qy 150 ATCAAAAAATTAGCTGGGTGTAGTGATGCTCTGTAGTCCAGTACTTCGGGAGGCTG 209
 Db 114902 ATCAAAACATTAGCCAGCGGTGTAGTGACGCTTATATCTTACCTAGTACTTTAGAGGCTG 114843
 Qy 210 AGCAAGAGATTGCTTTGAACCTGGGAGGAGAGTTTGCAGTGGAGGAGATCCACCAC 269
 Db 114842 AGGCACGAGATCCCTTTGAGCTGGGAGGAGAGTTTGCAGTGGAGGAGATCAACGCCAC 114783
 Qy 270 TGCATCCAGCTGGGCGACACAGCGAGACTCTATCTCAAAAAATAAATAAATAAATA 329
 Db 114782 TGCATTCAGCTGGGCGACAGCAAGATTTCCATCTCAAAAAATAAATAAATAAATA 114723
 Qy 330 AAGGTCGGAGAGAAACAAAACTAATAAGATTCTCTGAAGGTAAAGCAGAG-ATACGTAAT 388
 Db 114722 ATAAATGTCAAGGGATCATGAGCCATGGAGAGGAATAAGATATTATGATATATGACTTG 114663
 Qy 389 TATATGTAATAAAGTTTAAATGCAATTTAACTGTAACTTTATTTGTTTATTTGTTTATAA 448
 Db 114662 ATTCGCAATAAATAATTTTACATCATAGAAATGATATCTTTATTTAGTATTTGGCTTTAG 114603
 Qy 449 AAGTAAACAAGCCAAAGTAATGCACTTCAAACTCTACATAAATATCTATTTATTTATGGAAG 508
 Db 114602 AAGAACCATAATTTTAAAAACATGGTTATAGAAATAGGATGTAATATCTATAGTCTAACA 114543
 Qy 509 TGAAGGCATCTATAATCCTTACTACCCAAAGATAACAGTTACATATTTCTCCAGATTTT 568
 Db 114542 AACATATAGTTAAAGGAAGGACAGAGCTCTAGAAGATAAATCTGAAGTTGATGCT 114483
 Qy 569 TGGGGCATACACTAGCTATTTTATTTGGGAAATTTTCCATGTGCGAGCATACCTAATTT 628
 Db 114482 ACCTAACAGAGTAATCAACAAAACTAAAGACAGCAGCTGTGATGATGAGGGGACACA 114423
 Qy 629 TTCTAAATGCTATGTAGTATTCATTTAAGGATGTTTCCATAATTTTAAAAATFACATGCT 688
 Db 114422 TTGTGGGTGGGCTAAACCCCTCATCTTAAACAAAAAGGTAAACAGGTAATAAATAAATTTGGT 114363
 Qy 689 --TTAAAGTAGAGAACTAGTGTGGCATGTGGCTCAGCCCTGT-ATCCAGCAGCTTTG 745
 Db 114362 AATCAGAAATAAAAACTAGGCTGGGCGACGCTGCTCAGCAGCTGTAATCCAGCAGCTTTG 114303
 Qy 746 GGAGGCGGAGCAAAATGGATCACTTGAAGTCCGGAGTTTGAGACCAAGCTGGACAAACATG 805
 Db 114302 GGAGGCCAAGGCGAGGATCACTTGAAGCCAGAGTTTCGAGACCAAGCTGGCCACATG 114243
 Qy 806 ATGAAACACCTCTCTAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAA 865
 Db 114242 ACAAAACCCCATCTCTACTAAAAATAAATAAATAAATAAATAAATAAATAAATAAATAA 114183
 Qy 866 AGTCCAGCTACTCAGAGTCTCGAGGAGGATC 901
 Db 114182 AGTCCAGCTACTTGGAGGCTGAGGAGGACAATC 114147

RESULT 9

LOCUS ALL138896 265738 bp DNA linear HTG 30-JAN-2002
 DEFINITION Homo sapiens chromosome 10 clone RP11-133M13, 16 unordered pieces.
 ACCESSION AL138896
 VERSION AL138896.11 GI:14625510
 KEYWORDS HTG; HTGS PHASE1; HTGS_CANCELLED.
 SOURCE Homo sapiens (human)
 ORGANISM Homo sapiens
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini;
 Hominidae; Homo.

REFERENCE

1 Chapman, J.
Direct Submission
Submitted (29-JAN-2002) Wellcome Trust Sanger Institute, Hinxton,
Cambridgeshire, CB10 1SA, UK. E-mail enquiries:
humquerry@sanger.ac.uk Clone requests: clonerequest@sanger.ac.uk
On Jul 6, 2001 this sequence version replaced gi:14586010.
----- Genome Center
Center: Wellcome Trust Sanger Institute
Center code: SC
Web site: http://www.sanger.ac.uk
Contact: humquerry@sanger.ac.uk
----- Project Information
Center project name: bA133M13
----- Summary Statistics
Assembly program: XGAP4; version 4.5
Sequencing vector: plasmid; 108752; 100% of reads
Chemistry: Dye-terminator Big Dye; 100% of reads
Consensus quality: 262635 bases at least Q40
Consensus quality: 263809 bases at least Q30
Consensus quality: 264520 bases at least Q20
Insert size: 265238; sum-of-contigs
Insert size: 141679; 7.5% error; agarose-fp
Quality coverage: 10.63x in Q20 bases; sum-of-contigs Quality
coverage: 20.65x in Q20 bases; agarose-fp

COMMENT

* NOTE: This is a 'working draft' sequence. It currently
* consists of 16 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.

* 1 32499: contig of 32499 bp in length
* 32500 32999: gap of 100 bp
* 32600 34912: contig of 2313 bp in length
* 34913 35012: gap of 100 bp
* 35013 106818: contig of 71806 bp in length
* 106819 106918: gap of 100 bp
* 106919 125896: contig of 18968 bp in length
* 125897 125986: gap of 100 bp
* 125987 145316: contig of 19330 bp in length
* 145317 145416: gap of 100 bp
* 145417 153004: contig of 7588 bp in length
* 153005 153104: gap of 100 bp
* 153105 155300: contig of 2196 bp in length
* 155301 155400: gap of 100 bp
* 155401 185898: contig of 30498 bp in length
* 185899 185998: gap of 100 bp
* 185999 188966: contig of 2968 bp in length
* 188967 189066: gap of 100 bp
* 189067 192900: contig of 3834 bp in length
* 192901 193000: gap of 100 bp
* 193001 197749: contig of 4749 bp in length
* 197750 197849: gap of 100 bp
* 197850 214650: contig of 16801 bp in length
* 214651 214750: gap of 100 bp
* 214751 218266: contig of 3516 bp in length
* 218267 218366: gap of 100 bp
* 218367 255095: contig of 36729 bp in length
* 255096 255195: gap of 100 bp
* 255196 264279: contig of 9084 bp in length
* 264280 264379: gap of 100 bp
* 264380 266738: contig of 2359 bp in length.
* Location/Qualifiers
1. .266738

FEATURES
source

/organism="Homo sapiens"
/mol_type="genomic DNA"
/db_xref="taxon:9606"
/chromosome="10"
/clone="RP11-133M13"
/clone_lib="RPC1-11.1"

misc_feature 1. .32499
/note="assembly_fragment:02113
fragment_chain:1"
misc_feature 32600. .34912
/note="assembly_fragment:02238
fragment_chain:1"
misc_feature 35013. .106818
/note="assembly_fragment:07688
fragment_chain:2"
misc_feature 106919. .125886
/note="assembly_fragment:03826
fragment_chain:2"
misc_feature 125987. .145316
/note="assembly_fragment:07862
fragment_chain:3"
misc_feature 145417. .153004
/note="assembly_fragment:01625
fragment_chain:3"
misc_feature 153105. .155300
/note="assembly_fragment:01523"
misc_feature 155401. .185898
/note="assembly_fragment:01783"
misc_feature 185999. .188966
/note="assembly_fragment:03724"
misc_feature 189067. .192900
/note="assembly_fragment:05406"
misc_feature 193001. .197749
/note="assembly_fragment:05944"
misc_feature 197850. .214650
/note="assembly_fragment:06359"
misc_feature 214751. .218266
/note="assembly_fragment:06448"
misc_feature 218367. .255095
/note="assembly_fragment:07266"
misc_feature 255196. .264279
/note="assembly_fragment:07507"
misc_feature 264380. .266738
/note="assembly_fragment:07874"

ORIGIN
Query Match 37.4%; Score 337.2; DB 14; Length 266738;
Best Local Similarity 63.8%; Pred. No. 2.6e-58;
Matches 559; Conservative 0; Mismatches 313; Indels 4; Gaps 3;

QY 30 GGGCACAGTACTCAAGCCTTAACCAACAGTACTTTGGAACTCCAAAGTGGTGGTGCATCAC 89
Db 86293 GGGCGCAGTGCTCACACCTGTAAATCGCAGCATTTTGGGAGGCCAAGTGGTGGTGCATCAC 86234
QY 90 TTGAGTGCAGAAAGTTTCGAGACAGCAGCCTGGTCAATATGTTGAAACCCCTATCTCTACTAAAA 149
Db 86233 TTGAGTGCAGAGTTTGGACACAGCCTGGCCCAACATGTTGAAACCCCGTCTCTACTAAAA 86174
QY 150 ATACAAAAATTAGCTGGGTGTAGTGATGCATGCTGTAGTCCAGTACTCTCGGAGGCTG 209
Db 86173 ATACAAACATTAGCCAGCGCTGGTAGTCACGCGCTATAATCTCTACTTCTTAGGAGGCTG 86114
QY 210 AGGCAAGAGAAATGCTTGAACCTGGGAGCAGAGTTCAGTGCAGCCAGATCCCAACAC 269
Db 86113 AGGCACAGAAATCCCTTGGAGCCTGGGAGCAGAGTTCAGTGCAGCCAGATCCCAACAC 86054
QY 270 TGCAGTCCAGCCTGGGCGACACAGCAGACTCTATCTCAAAAAATAAATAATAATAATA 329
Db 86053 TCATTTCAGCCTGGGCGACAGCAGCAAGATTCATCTCAAAAAATAAATAATAATAATA 85994
QY 330 AAGGATCGGAGAGAAACAAACCTAATAAGATTCCTGAAAGGTAAAGCAGAG-ATACGTAAT 388
Db 85993 ATAAATGTCAGAGGATCATGAGCCATGGAAGAAATAGAAATATATATATGACTTGT 85934
QY 389 TATATGTAATAAGTTTAAATGCATTTTAACTGTAATCTTATGTTTATTTTGGTTATAA 448
Db 85933 ATTCTGCAATAAATATTTTACATCATAGAAATGATCTTTTATAGTATTTGGCTTTAG 85874
QY 449 AAGTAAACCAAGCCAAAGTAATGTAACCTTCAAACTCTACATAAATATCTATTATGGAAG 508

Db	85873	AAAGAACCAATATATTTTAAACAATGGTTATAGAAATAGGATGTAAATATCTATAGTCTAACA	85814
Qy	509	TGGAAGGCATCTATAATCTTACTACCCAAAGATAACCAAGTACATATCTCCAGATTTT	568
Db	85813	AACAGTATAGTTAAAGGAAGGCAAGAGCTCATGAAGATAAATCTGAAAGTTGATGCT	85754
Qy	569	TGGGGGCATACATAGCTTTTTTTTATTTGGGAAATTTCCATGTGCGAGGCATACCTAAATTT	628
Db	85753	ACCTAACAGAAGTAATCAACAAAAACAACCTAAAGACAGCAGTGTGATGATGAGGCGGCACA	85694
Qy	629	TTCTAAATGTCATGTAGTATTCATTTAAGGATGTTCCATTAATTTTTTAAATATACATGCT	688
Db	85693	TTGTGGGTGGGTAAACCCCTCATCTTTAAACAAAAGGTAAACAGGTAAATAACTAAAAATTGGT	85634
Qy	689	--TTAAAGTAGAGAAACTAGGTTGGGCATGTGGCTCAGCCCTGT-ATCCGAGCACTTTG	745
Db	85633	AATCAAGAAATAAANAATAGGCTGGCAGCGTGGTCAACCTGTAAATCCAGACATTTG	85574
Qy	746	GGAGGCGGAGGCAAAATGGATCACTTGAGTCCGGAGTTTGAACAGCAGCTGGCAACAATG	805
Db	85573	GGAGGCGAAGGCGAGGAGCATCACTTGAGGCGAGGAGTTTCGAGACAGCCTGGCCAAATG	85514
Qy	806	ATGAACACCCCTCTCTAATAAATAAATAAATAAATAGCTGGCATGCTGGCAGACCTGT	865
Db	85513	ACAAAACCCCATCTCTACTATAAATAAATAAATAAATAGCCGGCATGCTGGCAGCGCCTGT	85454
Qy	866	AGTCCCAGCTACTTCAGGAGTCTGAGGCGAGGATATC	901
Db	85453	AGTCCCAGCTACTTTGGAGGCTTGAGGCGAGGACATC	85418

RESULT 10
AP002020/c

LOCUS
DEFINITION
Homo sapiens chromosome 4 clone 2087A2 map 4q22-q24, *** SEQUENCING
IN PROGRESS ***, 7 unordered pieces.

ACCESSION
VERSION
KEYWORDS
SOURCE
ORGANISM

AP002020
AP002020.2 GI:9795241
HTG: HTGS_PHASE1.
Homo sapiens (human)
Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini;
Hominiidae; Homo.
Tsai,S.F.
Direct Submission
Submitted (09-MAY-2000) Shih-Feng Tsai, National Yang-Ming
University, Institute of Genetics, 155 Li-Rong St. Section 2,
Peitou, Taipei, Taiwan 11221, Republic of China
(E-mail: ymptsai@ym.edu.tw, URL: http://genome.ym.edu.tw/,
Tel: 886-2-28267043, Fax: 886-2-28264930)
On Aug 11, 2000 this sequence version replaced gi:7798580.
These sequences are draft human sequences, not finished sequences.
These sequences are unordered pieces. And gaps between the contigs
of the same clone are represented as 100 N.
* NOTE: This is a 'working draft' sequence. It currently
* consists of 7 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.

COMMENT

1 6503: contig of 6503 bp in length
* 6504 6603: gap of 100 bp
* 6604 15759: contig of 9156 bp in length
* 15760 15959: gap of 100 bp
* 15960 25981: contig of 10122 bp in length
* 25982 26081: gap of 100 bp
* 26082 38177: contig of 12096 bp in length
* 38178 38277: gap of 100 bp

Db 86597 AGTTCAAGACAGCCTGGCCACACGTGTGAACCCCTGCTCTACTATAAAATACAAAACCTT 86538

Oy 840 AGCTGGGCGATGGTGGCAAGCACACCTGTAGTCCAGCTACTCAGGAGTCTGAGGCGAGGAGTA 899

Db 86537 AGCGGGCGTGTGGTGGCGGCGCTGTAACTCTAGCTATTGGGAGGCTGAGGCGAGGAGAA 86478

Oy 900 TC 901

Db 86477 TC 86476

RESULT 11

AC126283/c

LOCUS

DEFINITION Homo sapiens BAC clone RP11-602N24 from 4, complete sequence.

ACCESSION AC126283

VERSION AC126283.3 GI:22091403

KEYWORDS HTG.

SOURCE Homo sapiens (human)

ORGANISM Homo sapiens

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini;

Hominae; Homo.

1 (bases 1 to 118642)

Sulston,J.E. and Waterston,R.

Toward a complete human genome sequence

Genome Res. 8 (11), 1097-1108 (1998)

9847074

2 (bases 1 to 118642)

Yoakum,M., Kozlowski,A. and Creason,K.

The sequence of Homo sapiens BAC clone RP11-602N24

Unpublished (2001)

3 (bases 1 to 118642)

Waterston,R.H.

Direct Submission

Submitted (04-JUL-2002) Genome Sequencing Center, Washington

University School of Medicine, 4444 Forest Park Parkway, St. Louis,

MO 63108, USA

4 (bases 1 to 118642)

Waterston,R.H.

Direct Submission

Submitted (29-JUL-2002) Genome Sequencing Center, Washington

University School of Medicine, 4444 Forest Park Parkway, St. Louis,

MO 63108, USA

5 (bases 1 to 118642)

Waterston,R.H.

Direct Submission

Submitted (02-AUG-2002) Genome Sequencing Center, Washington

University School of Medicine, 4444 Forest Park Parkway, St. Louis,

MO 63108, USA

6 (bases 1 to 118642)

Waterston,R.

Direct Submission

Submitted (10-AUG-2002) Department of Genetics, Washington

University, 4444 Forest Park Avenue, St. Louis, Missouri 63108, USA

On Aug 2, 2002 this sequence version replaced gi:22002234.

----- Genome Center

Center: Washington University Genome Sequencing Center

Center code: WUGSC

Web site: <http://genome.wustl.edu/gsc>

Contact: sapiens@wustl.edu

----- Summary Statistics

Center project name: H_NH0602N24

NOTICE: This sequence may not represent the entire insert of this

clone. It may be shorter because we only sequence overlapping

clone sections once, or longer because we provide a small overlap

between neighboring data submissions.

This sequence was finished as follows unless otherwise noted:

all regions were double stranded, sequenced with an alternate

chemistry, or covered by high quality data (i.e., phred quality >=

30): an attempt was made to resolve all sequencing problems, such

as compressions and repeats; all regions were covered by sequence

from more than one subclone; and the assembly was confirmed by

restriction digest.

MAPPING INFORMATION:

Mapping information for this clone was provided by Dr. John D.

McPherson, Department of Genetics, Washington University, St. Louis

MO. For additional information about the map position of this

sequence, see <http://genome.wustl.edu/gsc>

SOURCE INFORMATION:

The RPCI-11 human BAC library was made from the blood of one male

donor, as described by Osogawa,K., Woon,P.Y., Zhao,B., Frengen,E.,

Tateno,M., Catanese,J.J. and de Jong,P.J. (1998) An improved

approach for construction of bacterial artificial chromosome

libraries. Genomics 51:1-8. The clone may be obtained either from

Research Genetics, Inc. (<http://www.resgen.com>) or Pieter de Jong

and coworkers at <http://www.chori.org>

VECTOR: pBACe3.6

NEIGHBORING SEQUENCE INFORMATION:

The clone sequenced to the left is AC004067, 2000 bp overlap; the

clone sequenced to the right is AC005509, 2000 bp overlap.

Polymorphisms exist between AC004067 and AC126283.

FEATURES

source

1. .118642

/organism="Homo sapiens"

/mol_type="genomic DNA"

/db_xref="taxon:9606"

/chromosome="4"

/map="4"

/clone="RP11-602N24"

/clone_lib="RPCI-11"

56. .78

/rpt_family="AT_rich"

109. .196

/rpt_family="MIR"

310. .412

/rpt_family="L2"

694. .775

/rpt_family="Tip100"

776. .1073

/rpt_family="Alu"

1074. .1160

/rpt_family="Tip100"

1182. .1469

/rpt_family="Alu"

1679. .2085

/rpt_family="L2"

2086. .2390

/rpt_family="Alu"

2391. .2719

/rpt_family="L2"

2720. .3023

/rpt_family="Alu"

3043. .3078

/rpt_family="AT_rich"

3151. .3452

/rpt_family="Alu"

3674. .3708

/rpt_family="AT_rich"

3901. .4585

/rpt_family="ERV1"

4975. .5569

/rpt_family="L2"

5804. .6054

/rpt_family="ERV1"

6306. .6609

/rpt_family="Alu"

6685. .7017

/rpt_family="MaLR"

ORGANISM

Homo sapiens

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini; Hominidae; Homo.

REFERENCE

1 (bases 1 to 161306)

Birren,B., Linton,L., Nusbaum,C. and Lander,E.

Unpublished

2 (bases 1 to 161306)

Birren,B., Linton,L., Nusbaum,C., Lander,E., Allen,N., Anderson,S., Barna,N., Bastien,V., Boguslavskiy,L., Bouckghalter,B., Brown,A., Camarata,J., Campopiano,A., Chang,J., Choepel,Y., Colangelo,M., Collins,S., Collymore,A., Cooke,P., Dearellano,K., Dewar,K., Diaz,J.S., Dodge,S., Faro,S., Ferreira,P., FitzHugh,W., Gage,D., Galagan,J., Gardyna,S., Ginde,S., Goyette,M., Graham,L., Grand-Pierre,N., Hagos,B., Heaford,A., Horton,L., Hulme,W., Iliev,I., Johnson,R., Jones,C., Karatas,A., LaRocque,K., Lamazares,R., Landers,T., Lehoczy,J., Levine,R., Liu,G., MacLean,C., Macdonald,P., Marquis,N., Matthews,C., McCarthy,M., McEwan,P., McKernan,K., McPheeters,R., Meldrim,J., Meneus,L., Mihova,T., Mienga,V., Murphy,T., Naylor,J., Nguyen,C., Norbu,C., Norman,C.H., O'Connor,T., O'Donnell,P., O'Neil,D., Oliver,J., Peterson,K., Phunkhang,P., Pierre,N., Pollara,V., Raymond,C., Retta,R., Rieback,M., Riley,R., Rise,C., Rogov,P., Roman,J., Rosetti,M., Roy,A., Santos,R., Schauer,S., Stange-Thomann,N., Seaman,S., Severy,P., Sougnez,C., Spencer,B., Stange-Thomann,N., Stojanovic,N., Strauss,N., Strausman,A., Talamas,J., Talamas,J., Tesfaye,S., Theodore,J., Travers,M., Travis,N., Trigilio,J., Vassiliev,H., Viel,R., Vo,A., Wilson,B., Wu,X., Wyman,D., Ye,W.J., Young,G., Zainoun,J., Zembek,L., Zimmer,A. and Zody,M.

TITLE

Direct Submission

JOURNAL

Submitted (25-MAR-2001) Whitehead Institute/MIT Center for Genome Research, 320 Charles Street, Cambridge, MA 02141, USA

REFERENCE

3 (bases 1 to 161306)

Birren,B., Linton,L., Nusbaum,C., Lander,E., Ali,A., Allen,N., Anderson,S., Barna,N., Bastien,V., Boguslavskiy,L., Bouckghalter,B., Brown,A., Camarata,J., Campopiano,A., Chang,J., Chazaro,B., Choepel,Y., Colangelo,M., Collins,S., Collymore,A., Cook,A., Cooke,P., Dearellano,K., Dewar,K., Diaz,J.S., Dodge,S., Faro,S., Ferreira,P., FitzHugh,W., Gage,D., Galagan,J., Gardyna,S., Ginde,S., Gord,S., Goyette,M., Graham,L., Grand-Pierre,N., Hagos,B., Heaford,A., Horton,L., Hulme,W., Iliev,I., Johnson,R., Jones,C., Kamat,A., Karatas,A., Kells,C., LaRocque,K., Lamazares,R., Landers,T., Lehoczy,J., Levine,R., Liu,G., MacLean,C., Macdonald,P., Major,J., Marquis,N., Matthews,C., McCarthy,M., McKernan,K., McPheeters,R., Meldrim,J., Meneus,L., Mihova,T., Mienga,V., Murphy,T., Naylor,J., Nguyen,C., Norbu,C., Norman,C.H., O'Connor,T., O'Donnell,P., O'Neil,D., Oliver,J., Peterson,K., Phunkhang,P., Pierre,N., Pollara,V., Raymond,C., Retta,R., Rieback,M., Riley,R., Rise,C., Rogov,P., Roman,J., Rosetti,M., Roy,A., Santos,R., Schauer,S., Schupback,R., Seaman,S., Severy,P., Spencer,B., Stange-Thomann,N., Stojanovic,N., Strauss,N., Strausman,A., Talamas,J., Tesfaye,S., Theodore,J., Topham,K., Travers,M., Travis,N., Trigilio,J., Vassiliev,H., Viel,R., Vo,A., Wilson,B., Wu,X., Wyman,D., Ye,W.J., Young,G., Zainoun,J., Zembek,L., Zimmer,A. and Zody,M.

TITLE

Direct Submission

JOURNAL

Submitted (27-MAR-2003) Whitehead Institute/MIT Center for Genome Research, 320 Charles Street, Cambridge, MA 02141, USA

COMMENT

On Nov 28, 2001 this sequence version replaced gi:15706088.

All repeats were identified using RepeatMasker:

Smit, A.F.A. & Green, P. (1996-1997)

http://ftp.genome.washington.edu/RM/RepeatMasker.html

Center: Whitehead Institute/ MIT Center for Genome Research

Center code: WBIR

Web site: <http://www-seq.wi.mit.edu>

Contact: sequence_submissions@genome.wi.mit.edu

----- Project Information

Center project name: L11893

Center clone name: 83_J_16

* NOTE: This is a 'working draft' sequence. It currently

* consists of 4 contigs. The true order of the pieces

* is not known and their order in this sequence record is

* arbitrary. Gaps between the contigs are represented as

* runs of N, but the exact sizes of the gaps are unknown.

* This record will be updated with the finished sequence

* as soon as it is available and the accession number will

* be preserved.

* 1 24848: contig of 24848 bp in length

* 24849 24948: gap of 100 bp

* 24949 132812: contig of 107864 bp in length

* 132813 132912: gap of 100 bp

* 132913 135523: contig of 2611 bp in length

* 135524 135623: gap of 100 bp

* 135624 161306: contig of 25683 bp in length.

FEATURES

source

1. 161306

/organism="Homo sapiens"

/mol_type="genomic DNA"

/db_xref="taxon:9606"

/chromosome="15"

/map="15"

/clone="RP11-83J16"

/clone_lib="RPC1-11 Human Male BAC"

24849..24948

/estimated_length=100

132813..132912

/estimated_length=100

135524..135623

/estimated_length=100

ORIGIN

Query Match 36.08; Score 324.2; DB 14; Length 161306;

Best Local Similarity 62.7%; Pred. No. 1.2e-55;

Matches 564; Conservative 0; Mismatches 313; Indels 22; Gaps 3;

QY 4 TCATATTAAATATAAATAAACAACCTCAGTTGGGCACAGTCAAGCTGTAAGCTGATGATGCC 63

DB 105318 TCAATATGAATAAAGTTGTACACCGGGTGCAGTGCCTATGCTTTAATCCAGCACT 105377

QY 64 TTGGAAGTCCAAAGTGGGTGATCACTTGAGGTGAGAAGTTGAGACACGCGCTGTCAAT 123

DB 105378 TTGGAGGCTGAAGCAGGTGGATCACTCGAGGTGAGGAGTTCAAGACTAGGCTGACCAAC 105437

QY 124 ATGGTGAACCTATCTCTACTAAATAATACAAAAATAGCTGGGTGTAGTGTGATGCC 183

DB 105438 ATGGTGAACCTATCTCTACTAAATAATACAAAAATAGCTGGGTGTGATGCC 105497

QY 184 TGTAGTCCAGCTACTCCGGAGGCTGAGGCAAGAGAATTGCTTGAACCTGGGAGGAG 243

DB 105498 TGTATCCAGCTACTCAGAGGCTGAGGCAAGAGAATTGCTTGAACCTGGGAGGAG 105557

QY 244 GTTCAGTGAACCGAGATCCACACCTGCTGAGCTGGGCGACAGCGAGACTCTA 303

DB 105558 GTTCAGTGAACCGAGATTCACCACTCCAGCTGGGCGACAGCGAGACTCTCA 105617

QY 304 TCTCAAAAATAATAATAATAATAAAGGATCGGAGAGAAACAAACTAATAAGATTCC 363

DB 105618 TCTCAAAAATAATAATAATAATAAAGGATCGGAGAGAAACAAACTAATAAGATTCC 105671

QY 364 TGAAGGTGAACGAGATACGTAATAATATATATATATATATATATATATATATATAT 423

DB 105672 -----TGTAAATGACATATCTTAAATATGATACATCTCTTCAGTGAA 105718

QY 424 ATCTATTGTTTATTTGTTTATAAAAGTAAACAGCCAAAGTATGCAACTTCAAACT 483

DB 105719 GATAGTAAATATATAGAAAAAATCATGATCTAAACAATATGATGTTTAACTTTTGG 105778

QY 484 CTACATAATATCTATTATGGAAGTGGAGGCACTCTAATACTCTACTACCAAGATAA 543

DB 105779 ATCATAAATTTGAATAGTAGAAGATCCAAAGCTGGGGAATCTTTAGGAGGAGGA 105838

QY 544 CCAGTTACATATTCCTCCAGATTTTGGGGCATACACTAGCTTTTATTTTGGGAAAT 603

|||||

Db 105839 AACAAACATAATGTTGTAGCACTATACATTTTAAATAATTTCTTTTGTGGAACATA 105898
QY 604 TTCCATGTCGAGCATACCTAAATTTTCTAAATGCTATGTAGTATTCCTAATTAAGGATG 663
Db 105899 TTTAAATTTAGGAATAAAATTTAGCCACAGACCTTAAATAGGCACAGGATCAAGTTA 105958
QY 664 TTCCATAATTTTAAATAACATGCTTTAAAGTAGAGAACTAGGTTGGGCATGGTGGCTC 723
Db 105959 TCCCTCAACTT--ATTTCCTATTAGAAATTCCTGAAACAGAGTTGGGCACGGTGATTC 106016
QY 724 AGCCTGTATCCAGCACTTTGGAGGCGGCGCAATGATGACTGTAGGTCGGAGTT 783
Db 106017 ATGCTGTAAACCAACACTTTGGAGGCTGTAGGCTGGTCAATCATTGAGGCGGAGGTT 106076
QY 784 TGAGACCAAGCTGGCAACATGATGAACACACCTTC-TCTAATAAAAAATACAAAAATTAGC 842
Db 106077 CAAGACCAAGCTGGCGCAACGTCGCGAACTCTATCTTCTACTCAAAATACAAAATTAGC 106136
QY 843 TGGGATGTTGGCAAGCACCTGTAGTCCAGTACTCAGGAGTCTTGAGGCGGAGGATTC 901
Db 106137 TGGGAGTGGTGGCACACATCTGTAATACCACTACTCGGGAAGATGAGCGCACGAGATC 106195

RESULT 13

AC018868
LOCUS AC018868 162100 bp DNA linear PRI 09-JAN-2002
DEFINITION Homo sapiens BAC clone RP11-83J16 from 15, complete sequence.
ACCESSION AC018868
VERSION AC018868.4 GI:13443246
KEYWORDS HTG.
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini;
Hominidae; Homo.
REFERENCE 1 (bases 1 to 162100)
AUTHORS Sulston, J.E. and Waterston, R.
TITLE Toward a complete human genome sequence
JOURNAL Genome Res. 8 (11), 1097-1108 (1998)
PUBMED 9847074
REFERENCE 2 (bases 1 to 162100)
AUTHORS Grewal, N., Hawkins, M., Maupin, R., Gregory, S. and Scherger, E.
TITLE The sequence of Homo sapiens BAC clone RP11-83J16
JOURNAL Unpublished (2001)
REFERENCE 3 (bases 1 to 162100)
AUTHORS Waterston, R.H.
TITLE Direct Submission
JOURNAL Submitted (21-DEC-1999) Genome Sequencing Center, Washington University School of Medicine, 4444 Forest Park Parkway, St. Louis, MO 63108, USA
REFERENCE 4 (bases 1 to 162100)
AUTHORS Waterston, R.H.
TITLE Direct Submission
JOURNAL Submitted (24-MAR-2001) Genome Sequencing Center, Washington University School of Medicine, 4444 Forest Park Parkway, St. Louis, MO 63108, USA
REFERENCE 5 (bases 1 to 162100)
AUTHORS Waterston, R.
TITLE Direct Submission
JOURNAL Submitted (07-NOV-2001) Department of Genetics, Washington University, 4444 Forest Park Avenue, St. Louis, Missouri 63108, USA
REFERENCE 6 (bases 1 to 162100)
AUTHORS Waterston, R.
TITLE Direct Submission
JOURNAL Submitted (09-JAN-2002) Department of Genetics, Washington University, 4444 Forest Park Avenue, St. Louis, Missouri 63108, USA
COMMENT On Mar 24, 2001 this sequence version replaced gi:7630827.
----- Genome Center
Center: Washington University Genome Sequencing Center
Center code: WUGSC
Web site: <http://genome.wustl.edu/gsc>
Contact: sapiens@watson.wustl.edu
----- Summary Statistics

Center project name: H_NH0083J16

NOTICE: This sequence may not represent the entire insert of this clone. It may be shorter because we only sequence overlapping clone sections once, or longer because we provide a small overlap between neighboring data submissions.

This sequence was finished as follows unless otherwise noted: all regions were double stranded, sequenced with an alternate chemistry, or covered by high quality data (i.e., phred quality >= 30); an attempt was made to resolve all sequencing problems, such as compressions and repeats; all regions were covered by sequence from more than one subclone; and the assembly was confirmed by restriction digest.

MAPPING INFORMATION:

Mapping information for this clone was provided by Dr. John D. McPherson, Department of Genetics, Washington University, St. Louis MO. For additional information about the map position of this sequence, see <http://genome.wustl.edu/gsc>

SOURCE INFORMATION:

The RPC1-11 human BAC library was made from the blood of one male donor, as described by Osoegawa, K., Woon, P.Y., Zhao, B., Frengen, E., Tateno, M., Catanese, J.J. and de Jong, P.J. (1998) An improved approach for construction of bacterial artificial chromosome libraries. Genomics 51:1-8. The clone may be obtained either from Research Genetics, Inc. (<http://www.resgen.com>) or Pieter de Jong and coworkers at the Roswell Park Cancer Institute (<http://bacpac.med.buffalo.edu>)

VECTOR: pBACE3.6

NEIGHBORING SEQUENCE INFORMATION:

The clone sequenced to the left is RP11-814P5. Actual start of this clone is at base position 1 of RP11-83J16; actual end is at base position 162100 of RP11-83J16.

The sequence from base position 136368 to 136447 was derived from PCR product of RP11-83J16 BAC DNA.

FEATURES	source
repeat_region	1. 162100 /organism="Homo sapiens" /mol_type="genomic DNA" /db_xref="taxon:9606" /chromosome="15" /map="15"
repeat_region	31. 332 /clone="RP11-83J16" /clone_lib="RPC1-11" /rpt_family="L1"
repeat_region	268. 309 /rpt_family="(GA)n"
repeat_region	852. 1141 /rpt_family="L1"
repeat_region	1142. 1430 /rpt_family="Alu"
repeat_region	1431. 1594 /rpt_family="L1"
repeat_region	1886. 3527 /rpt_family="L1"
repeat_region	2849. 2877 /rpt_family="AT-rich"
repeat_region	3598. 4454 /rpt_family="L1"
repeat_region	4595. 4618 /rpt_family="(A)n"
repeat_region	6431. 6622 /rpt_family="L2"
repeat_region	6923. 7015 /rpt_family="L2"
misc_feature	7129. 7149 /notes="similar to Homo sapiens EST BE560643 (NID:9804363)"

```
repeat_region 7131..7356 /rpt_family="Alu"
repeat_region 7295..7322 /rpt_family="AT_rich"
repeat_region 7473..7582 /rpt_family="MERL_type"
misc_feature 7504..7992 /note="similar to Sus scrofa EST BG381987 (NID:g13306459)"
repeat_region 8013..8164 /rpt_family="MERL_type"
repeat_region 8678..8882 /rpt_family="MIR"
repeat_region 8912..9211 /rpt_family="Alu"
repeat_region 9031..9061 /rpt_family="AT_rich"
repeat_region 9195..9215 /rpt_family="AT_rich"
repeat_region 9212..9312 /rpt_family="MIR"
repeat_region 9653..9695 /rpt_family="MIR"
repeat_region 9768..9791 /rpt_family="CTATT)n"
repeat_region 9833..10006 /rpt_family="AT_rich"
repeat_region 10012..10133 /rpt_family="MERL_type"
misc_feature 10065..10079 /note="similar to Homo sapiens EST BE278798 (NID:g9153784)"
repeat_region 10275..10369 /rpt_family="L2"
repeat_region 11854..11931 /rpt_family="ERV1"
repeat_region 11925..12288 /rpt_family="ERV1"
repeat_region 13262..13418 /rpt_family="Alu"
repeat_region 13735..13915 /rpt_family="L1"
repeat_region 13893..13924 /rpt_family="AT_rich"
misc_feature 14140..14156 /note="similar to Homo sapiens EST AA808895 (NID:g2878301) nw1606.81"
repeat_region 14195..14217 /rpt_family="AT_rich"
repeat_region 14646..14709 /rpt_family="CT-rich"
repeat_region 14721..14758 /rpt_family="(CAT)n"
repeat_region 15600..15775 /rpt_family="L1"
repeat_region 15884..16117 /rpt_family="L1"
repeat_region 16118..16241 /rpt_family="Alu"
repeat_region 16242..16907 /rpt_family="L1"
repeat_region 16547..16574 /rpt_family="AT_rich"
repeat_region 16908..17191 /rpt_family="Alu"
repeat_region 17192..17844 /rpt_family="L1"
repeat_region 17856..17975 /rpt_family="Alu"
repeat_region 18010..18814 /rpt_family="L1"
repeat_region 18928..19231 /rpt_family="Alu"
repeat_region 19210..19231
```

```
repeat_region /rpt_family="(A)n"
19412..19772 /rpt_family="WalR"
repeat_region 19960..21700 /rpt_family="L1"
repeat_region 21702..22516

Query Match 36.0%; Score 324.2; DB 8; Length 162100;
Best Local Similarity 62.7%; Pred. No. 1.2e-55;
Matches 564; Conservative 0; Mismatches 313; Indels 22; Gaps 3;

Qy 4 TCATATTAAATTAATAAATACTCAGTTGGGCACAGTACTCAAGCCTGTAACCAAGTACT 63
Db 105325 TCAATATGAATAAATAGTGTACACCGGGTGCACTGCTCATGCTTAAATCCAGCACT 105384
Qy 64 TTGGAAGTCCAAAGTGGGTGATCCTTGAGGTGAGAGTTTCGAGACAGCCTGGTCAAT 123
Db 105385 TTGGAGGCTGAAGCAGGTGGATCACCTGAGGTGAGAGTTTCAAGATAGCTGACCAAC 105444
Qy 124 ATGGTGAACCCCTATCTCTACTAAAAATACAAAAATTAGCTGGGTGTAGTGATGATGCC 183
Db 105445 ATGGTGAACCCCACTCTCTCTAAAAATACAAAAATTAGCTGGGTGTAGTGATGCC 105504
Qy 184 TGTAGTCCCACTACTCGGAGGCTGAGGCAAGAGAATTGCTTGAACCTGGGAGGCAAG 243
Db 105505 TGTAAATCCCACTACTCAGGAGGCTGAGGCAAGAGAATTGCTTGAACCCGGAGGCGAA 105564
Qy 244 GTTGAGTGAGCCAGATCCCACTGCACTCCAGCTGGGCGACAGCAGGAGACTCTA 303
Db 105565 GTTGAGTGAACCCAGATTCACCATTTGCACTCCAGCTGGGCGACAGGAGACTCCA 105624
Qy 304 TCTCAAAAAATAAATAAATAAAGGATCGGAGAGAAACAAAATAATAAGATTCC 363
Db 105625 TCTCAAAAAATAAATAAAGAGAAAGAAAGAAATAAGTGTATTCTAATAA----- 105678
Qy 364 TGAAGGTAAGCAGAGATACGTAATAATATATCTAATAAAGTTTAAATGTCATTTTAACTGTA 423
Db 105679 -----TGCTAAAAATGACATATCTTAAAAATGTATCACAATCTCTTCAAGTAA 105725
Qy 424 ATCTATTGTTTATTTGTTTATAAAGTAACAGCCAAAAGTAATGCAACTTCAAACT 483
Db 105726 GATAGTATAATTTATAGAAAAAATCATGATATCTAAACAATATGATGTTTAACTTTTG 105785
Qy 484 CTACATAATATCTATTATGGAAGTGAAGCATCTATAATCTCTACTACCCAAAGATAA 543
Db 105786 ATCAATAATTGAATAGTAGAAGATCCAAAGCCTGGGAAATCCTTTAGGGAAGGAGGA 105845
Qy 544 CCAGTTACATATTCCTCCAGATTTTGGGGCATACACTAGCTTTTTTTTATTTGGGAAAT 603
Db 105846 AACAAAAACATAATGTTGTAGCACTATACATTTTAAAAAATTTATTTTCTTTTGTGAACTA 105905
Qy 604 TTCCATGTGAGGCACTACCTAATTTTCTAATGTCTATGTAGTATTCATTTAAGGATG 663
Db 105906 TTTAAATTTTAGGAATAAAATTTCTAGCCACAAGACCTTTAATAGGCACAGGATCAACGTTA 105965
Qy 664 TTCCATAATTTTAAATAACATGCTTTTAAAGTAGAAGAACTAGGTTGGGCATGCTGGCTC 723
Db 105966 TCTTCAAACTT--ATTTCCATTAGAAGAAATTCCTGAAAAACAGGTTGGGCACGGTGATTC 106023
Qy 724 ACGCCTGTATCCAGCACTTTTGGGAGGCGGAGGCAAAATGGATCATCTGAGGTCGGAGTT 783
Db 106024 ATGCGCTGTAAACCCACACTTTTGGGAGGCTGAGGTGGGTCAATCACTTGGGCGCAGGAT 106083
Qy 784 TGAGACCGCCTGGAGCAACATGATGAACACCCCTC-TCTAATAAATAACAAAAATTAGC 842
Db 106084 CAAGACCGCCTGGCCCAACGTCGGCAAACTCTATCTTACTCAAAATCAAAAAATTAGC 106143
Qy 843 TGGGCATGTGCAAGCAGCCTGTAGTCCAGCTACTCAGGAGTCTGAGGACAGGATATC 901
Db 106144 TGGGAGTGGTGCAACACATCTGTATATACCAGTCTACTCGGGAAGATGAGGCACGAGAATC 106202

RESULT 14
```

```

AC090783/c
LOCUS       AC090783               180230 bp      DNA      linear      HTG 30-JUL-2001
DEFINITION  Homo sapiens chromosome 15 clone RP11-205B5 map 15, WORKING DRAFT
SEQUENCE, 11 unordered pieces.
AC090783
AC090783.2  GI:15028544
VERSION     HTG; HTGS_PHASE1; HTGS_DRAFT; HTGS_FULLTOP.
KEYWORDS    Homo sapiens (human)
SOURCE      Homo sapiens
ORGANISM    Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini;
            Homidae; Homo.
REFERENCE   1 (bases 1 to 180230)
AUTHORS    Birren,B., Linton,L., Nusbaum,C., Lander,E., Allen,N., Anderson,S.,
            Camarata,J., Campotiano,A., Choepel,Y., Colangelo,M., Collins,S.,
            Collymore,A., Cooke,P., DeArellano,K., Dewar,K., Diaz,J.S.,
            Dodge,S., Faro,S., Ferreira,P., FitzHugh,W., Gage,D., Galagan,J.,
            Gardyna,S., Ginde,S., Goyette,M., Graham,L., Grand-Pierre,N.,
            Hagos,B., Heaford,A., Horton,L., Hulme,W., Iliev,I., Johnson,R.,
            Jones,C., Karas,A., Lacroque,K., Lamazares,R., Landers,T.,
            Lehoczy,J., Levine,R., Liu,G., Maclean,C., Macdonald,P.,
            Marquis,N., Matthews,C., McCarthy,M., McEwan,P., McKernan,K.,
            McPheeters,R., Meldrim,J., Meneus,L., Mihova,T., Mlenga,V.,
            Murphy,T., Naylor,J., Nguyen,C., Norbu,C., Norman,C.H.,
            O'Connor,T., O'Donnell,P., O'Neil,D., Oliver,J., Peterson,K.,
            Phunkhang,P., Pierre,N., Pollara,V., Raymond,C., Retta,R.,
            Rieback,M., Riley,R., Rise,C., Rogov,P., Roman,J., Rosetti,M.,
            Roy,A., Santos,R., Schauer,S., Schupback,R., Seaman,S., Severy,P.,
            Sounguez,C., Spencer,B., Stange-Thomann,N., Stojanovic,N.,
            Strauss,N., Subramanian,A., Talamas,J., Tesfaye,S., Theodore,J.,
            Travers,M., Travis,N., Triglio,J., Vassiliev,H., Viel,R., Vo,A.,
            Wilson,B., Wu,X., Wyman,D., Ye,W.J., Young,G., Zainoun,J.,
            Zembek,L., Zimmer,A. and Zody,M.
DIRECT SUBMISSION
TITLE       Submitted (10-MAR-2001) Whitehead Institute/MIT Center for Genome
JOURNAL     Research, 320 Charles Street, Cambridge, MA 02141, USA
COMMENT    On Jul 29, 2001 this sequence version replaced gi:13270652.
            All repeats were identified using RepeatMasker:
            Smit, A.F.A. & Green, P. (1996-1997)
            http://ftp.genome.washington.edu/RM/RepeatMasker.html
            ----- Genome Center
            Center: Whitehead Institute/ MIT Center for Genome Research
            Center code: WIBR
            Web site: http://www-seq.wi.mit.edu
            Contact: sequence_submissions@genome.wi.mit.edu
            ----- Project Information
            Center project name: L12756
            Center clone name: 205_B_5
            ----- Summary Statistics
            Sequencing vector: Plasmid; n/a; 100% of reads
            Chemistry: Dye-terminator Big Dye; 100% of reads
            Assembly program: Phrap; version 0.960731
            Consensus quality: 176672 bases at least Q40
            Consensus quality: 177894 bases at least Q30
            Consensus quality: 178437 bases at least Q20
            Insert size: 176000; agarose-fp
            Insert size: 179230; sum-of-contigs
            Quality coverage: 6.8 in Q20 bases; agarose-fp
            Quality coverage: 6.7 in Q20 bases; sum-of-contigs
            -----
            * NOTE: This is a 'working draft' sequence. It currently
            * consists of 11 contigs. The true order of the pieces
            * is not known and their order in this sequence record is
            * arbitrary. Gaps between the contigs are represented as
            * runs of N, but the exact sizes of the gaps are unknown.
            * This record will be updated with the finished sequence
            * as soon as it is available and the accession number will
            * be preserved.
FEATURES             Location/Qualifiers
     source            1..180230
                        /organism="Homo sapiens"
                        /mol_type="genomic DNA"
                        /db_xref="taxon:9606"
                        /chromosome="15"
                        /map="15"
                        /clone="RP11-205B5"
                        /clone_lib="RPC1-11 Human Male BAC"
     misc_feature       1..63929
                        /note="assembly_fragment"
                        clone_end:SP6
                        vector_side:left
     gap                63930..64029
                        /estimated_length=100
     misc_feature       64030..66793
                        /note="assembly_fragment"
     gap                66794..66893
                        /estimated_length=100
     misc_feature       66894..69526
                        /note="assembly_fragment"
     gap                69527..69626
                        /estimated_length=100
     misc_feature       69627..72423
                        /note="assembly_fragment"
     gap                72424..72523
                        /estimated_length=100
     misc_feature       72524..76915
                        /note="assembly_fragment"
     gap                76916..77015
                        /estimated_length=100
     misc_feature       77016..81564
                        /note="assembly_fragment"
     gap                81565..81664
                        /estimated_length=100
     misc_feature       81665..94020
                        /note="assembly_fragment"
     gap                94021..94120
                        /estimated_length=100
     misc_feature       94121..107816
                        /note="assembly_fragment"
     gap                107817..107916
                        /estimated_length=100
     misc_feature       107917..136957
                        /note="assembly_fragment"
     gap                136958..137057
                        /estimated_length=100
     misc_feature       137058..158034
                        /note="assembly_fragment"
     gap                168035..168134
                        /estimated_length=100
     misc_feature       168135..180230
                        /estimated_length=100

```

ORIGIN		/note="assembly_fragment clone_end:T7 vector_side:right"		Query Match Best Local Similarity Matches 564; Conservative		36.0%; 62.7%; 0;		Score 324.2; DB 14; Pred. No. 1.2e-55; Mismatchches 313;		Length 180230; Indels 22; Gaps 3;			
Qy	4	TCATATTAAATTAATAAATAACTCAGTTGGGCACAGTGACTCAAGCCTGTAAACACACAGTACT	63										
Db	164565	TCAAATATGAATAATAGTTTACACCGGTGCACTGCTGAGGTGAGGAGTTCAAGACTAGCTTCAAGCACT	164506										
Qy	64	TTGGAAGTCCAAAGTGGGTGGATCACTTTGAGGTGAGAAATTCGAGACAGAGCTGTGTAAT	123										
Db	164505	TTGGAGGCTGAAGCAGGTGGATCACTCAGGTGAGGAGTTCAAGACTAGCTTCAAGCACT	164446										
Qy	124	ATGGTGAACCCCTATCTCTACTAATAAATACAAAATTAAGCTGGGTGATGATGATGCC	183										
Db	164445	ATGGTGAACCCCATCTCTACTAATAAATACAAAATTAAGCTGGGTGATGATGATGCC	164386										
Qy	184	TGTAGTCCCAAGTACTCGGGAGGCTGAGGCAAGAGAAATTCCTTGAACCTGGGAGGACAG	243										
Db	164385	TGTAATCCCAAGCTACTCAGGAGGCTGAGGCAAGAGAAATTCCTTGAACCCGGAGGCGAA	164326										
Qy	244	GTTCAGTGAAGCCGAGATCCCACTGCACTCCAGCTGGGCGACACAGCGAGACTCTTA	303										
Db	164325	GTTCAGTGAACCGAGATTCACCATTCGACTCCAGCTGGGCGACAGAGTGAAGTCCCA	164266										
Qy	304	TCTCAAAAATAAATAAATAAATAGGATCGGAGAGAAACAAACTAATAAGATTCC	363										
Db	164265	TCTCAAAAATAAATAAATAAATAGGATCGGAGAGAAACAAACTAATAAGATTCC	164212										
Qy	364	TGAAGGTGAAGCAGAGATACGTAATTTATATGTAATAAGTTTAAATGCACTTTTAACTGTA	423										
Db	164211	-----TGCTAAATGACATATTCCTTAAATGTATCATCTCTCTCACTGTA	164165										
Qy	424	ATCTTATTTGTTTATTTGTTTATAAAGTAAACCAAGCCAAAAGTAAATGCAACTTCAAACT	483										
Db	164164	GATAGTAAATTTATAGAAAATTCATGATCTTAACATAATGATGTTTAACTTTTGTG	164105										
Qy	484	CTACATAAATATCTATTATGGAAGTGGAGGCACTATAATCCTACTACCCAAAGATAA	543										
Db	164104	ATCATAAATTTGAATAGTAGAAGAATCCAAAGCCTGGGGAATCCTTTAGGGAAGGAGGA	164045										
Qy	544	CCAGTTACATATTCCTCCAGATTTTGGGCGATACACTAGCTTTTATTTGGGAAAT	603										
Db	164044	AACAAAACATAATGTTGTAGCACTATACATTTTAAATAATATTTCTTTGTGGAACTA	163985										
Qy	604	TTCCATGTGAGGCATACCTTAATTTTCTAAATGTCTATGATGATATCCATTTAAGGATG	663										
Db	163984	TTTAAATTTTAGGAATAAAATTCCTAGCCCAAGACCTTAATAGGCAAGGATCAACGTTA	163925										
Qy	664	TTCCATAATTTTAAATACATGCTTTTAAAGTAGAAGAACTAGTTGGGCATGTGGCTC	723										
Db	163924	TCCTCAAACTT--ATTTTCCATTAGAAGAAATTCCTGAAACAGGTTGGGCACGGTATTC	163867										
Qy	724	ACGCCTGTATCCAGCACTTTTGGGAGCGCCAGGCAAAATGGATCACTTTGAGGTCGGAGTT	783										
Db	163866	ATGCTGTATACCCCAACATTTTGGGAGGCTGAGGTGGGTCAATCACTTTGAGCCAGGAGTT	163807										
Qy	784	TGAGACAGCCTGGACAACTATGTAACACCTC--TCTAATAAATAACAAAATTAGC	842										
Db	163806	CAAGACAGCCTGGCCAAACGTCGGGAAACTCTATCTTCTACTCAAAATACAAAATTAGC	163747										
Qy	843	TGGGCATGTGGCAAGCACTTGTAGTCCAGCTACTCAGGAGTCTGAGGCGAGGATC	901										
Db	163746	TGGGAGTGTGGCACACATCTGTAATACCAGCTACTCGGGAAGATGAGGCAACGAGATC	163688										
RESULT 15													
AL591215/c													
LOCUS													
AL591215													
154592 bp DNA linear PRI 18-MAY-2005													

DEFINITION	Human DNA sequence from clone RP11-20D12 on chromosome 9 Contains the 3' end of a novel gene, possible ortholog of mouse RIKEN CDNA 170028Pi4 gene, complete sequence.
ACCESSION	AL591215 AC025064
VERSION	AL591215.12 GI:15908225
KEYWORDS	HTG; 170028Pi4.
SOURCE	Homo sapiens (human)
ORGANISM	Homo sapiens
REFERENCE	1 (bases 1 to 154592)
AUTHORS	Garnier, P.
TITLE	Direct Submission
JOURNAL	Submitted (13-MAY-2005) Wellcome Trust Sanger Institute, Hinxton, Cambridgeshire, CB10 1SA, UK. E-mail enquiries: vegas@sanger.ac.uk
COMMENT	Clone requests: clonerequest@sanger.ac.uk On Sep 28, 2001 this sequence version replaced gi:15723808. The following abbreviations are used to associate primary accession numbers given in the feature table with their source databases: Em., EMBL; Sw., SWISSPROT; Tr., TREMBL; Wp., WORMPEP; Information on the WORMPEP database can be found at http://www.sanger.ac.uk/Projects/C_elegans/wormpep This sequence was generated from part of bacterial clone contigs of human chromosome 9, constructed by the Sanger Centre Chromosome 9 Mapping Group Further information can be found at http://www.sanger.ac.uk/HGP/Chr9 RP11-20D12 is from the library RPCI-11.1 constructed by the group of Pieter de Jong. For further details see http://www.chori.org/bacpac/home.htm VECTOR: pBACe3.6 ----- Genome Center Center: Wellcome Trust Sanger Institute Center code: SC Web site: http://www.sanger.ac.uk Contact: vegas@sanger.ac.uk ----- Draft Sequence Produced by Whitehead Institute/MIT Center for Genome Research, 320 Charles Street, Cambridge, MA 02141, USA http://www-seq.wi.mit.edu This sequence was finished as follows unless otherwise noted: all regions were either double-stranded or sequenced with an alternate chemistry or covered by high quality data (i.e., phred quality >= 30); an attempt was made to resolve all sequencing problems, such as compressions and repeats; all regions were covered by at least one subclone; and the assembly was confirmed by restriction digest, except on the rare occasion of the clone being a YAC. Location/Qualifiers 1..154592 /organism="Homo sapiens" /mol_type="genomic DNA" /db_xref="taxon:9606" /chromosome="9" /clone="RP11-20D12" /clone_lib="RPCI-11.1" 1 /note="Clone left end: RP11-20D12" Join(AL162412.11:177963..178201,3018..3126,15056..15174,.16410..16478,45339..45396,64455..64733) /locus_tag="RP11-20D12.1-001" Join(AL162412.11:177963..178201,3018..3126,15056..15174,.16410..16478,45339..45396,64455..64733) /locus_tag="RP11-20D12.1-001" /note="match: ESTs: AA416734.1 AA416815.1 AA757101.1 AF168674.1 AW654647.1 BB573244.1 BF059207.1 BG193301.1 BG194901.1 BU944256.1 BU961648.1 match: CDNAS: AK006473.1 AK014972.1" Join(AL162412.11:178050..178201,3018..3126,15056..15174,.16410..16478,45339..45396,64455..64637) /locus_tag="RP11-20D12.1-001" /standard_name="OTTHUMP0000021439" /note="match: proteins: Q9CCQ3"

```
/codon_start=1
/db_protein_id="CAH71316.1"
/db_xref="GI:55665830"
/db_xref="UniProt/TREMBL:Q5VTT2"
/translation="MDSLDKSCDMDKQHWLEIGPDLVERKGSILTRSHHKYSK
PVLVYSHRDREAPKGYDVBGPKEVKKLCNSTYRLGTDSPWTSFKLSQML
NTEWEMKSKALLNEETVSSGIIERVTLGPATGFAVFRPHPPDMSKMCALTYSSEY
VPFYDYQHPAQCDQDSIVHRKCRSQFTDLNGSKRFGINTWHDSESGIYANSVDVKQL
YPLTSGPIVPI"
gene      join(AL162412.11:178130.178201,3018.3126,15056.15174,
16410.16659)
/m locus_tag="RP11-20D12.1-002"
/m locus_tag="RP11-20D12.1-002"
/m note="match: ESTs: AA431486.1 AA948291.1 AW014133.1
BE502978.1 BG192998.1 B1825449.1"
gene      join(AL162412.11:178090.178201,15056.15174,16410.16680)
/m locus_tag="RP11-20D12.1-003"
/m locus_tag="RP11-20D12.1-003"
/m note="match: ESTs: BG194901.1"
polyA_signal      16661.16666
polyA_site      /locus_tag="RP11-20D12.1-003"
16680
gene      /locus_tag="RP11-20D12.1-003"
join(43556.43623,45339.45396,64455.64728)
/m locus_tag="RP11-20D12.1-004"
mRNA      join(43556.43623,45339.45396,64455.64728)
/locus_tag="RP11-20D12.1-004"
/note="match: ESTs: BG194901.1"
polyA_signal      64707.64712
polyA_site      /locus_tag="RP11-20D12.1-004"
64728
misc_feature      /locus_tag="RP11-20D12.1-004"
126192
/misc_feature      /note="Clone_right_end: RP11-385E2"
129644
/misc_feature      /note="Clone_left_end: RP11-44J18"
154592
/misc_feature      /note="Clone_right_end: RP11-20D12"
154592
ORIGIN
Query Match      35.9%; Score 323.8; DB 8; Length 154592;
Best Local Similarity      62.2%; Pred. No. 1.4e-55;
Matches 564; Conservative 0; Mismatches 332; Indels 11; Gaps 3;
Qy      4 TCATATTAAATAAATACTCAGTTGGGCAAGTCAAGCTGTAACCAAGTACT 63
Db      117733 TAAATAAGAGTAATAATACGGCTGGGCAAGTCAAGCTGTAATCCAGCACT 117674
Qy      64 TTGGAAGTCCAGGTGGGTGATCAGTGGTGAAGTTCGAGACCGAGCTGTCAT 123
Db      117673 TTGGAGGTGGAGCGGCGAGATCAATTTGAGGACAAAGATTTCAAGACCGAGCTGGGCAAC 117614
Qy      124 ATGGTGAACCTCTACTCTACTAAAAATACAAAAATAGCTGGGTGATGATCATGCC 183
Db      117613 ATGACGAAACCTGCTCTACTTGAATAACAAAAATAGCTGGGTGATGATCATGCC 117554
Qy      184 TGTAATCCAGCTGCTGGGAGGCTGAGGCAAGAAATTCCTGAACTGGGAGCGAG 243
Db      117553 TGTAATCCAGCTGCTGGGAGGCTGAGGCAAGAAATTCCTGAACTGGGAGCGAG 117494
Qy      244 GTTCGAGTCCAGGATCCACCACTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 303
Db      117493 GTTCGAGTCCAGGATCCACCACTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 117434
Qy      304 TCTCAAAAAATAAATAAATAAATAA-----GGATCGGAGAGAAACAAACTAA 354
Db      117433 TCTCAAAAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAA 117374
Qy      355 TAAGATTCTGAGGTGAGGAGAGATACGTAAATATATATGTAATAAAGTTTAAATGCATT 414
Db      117374 TAAGATTCTGAGGTGAGGAGAGATACGTAAATATATATGTAATAAAGTTTAAATGCATT 414
```

Search completed: February 9, 2006, 20:42:30
Job time : 4495.31 secs

THIS PAGE BLANK (USPTO)

GenCore version 5.1.1.7
Copyright (c) 1993 - 2006 Bioacceleration Ltd.

OM nucleic - nucleic search, using sw model

Run on: February 9, 2006, 14:19:56 ; Search time 580.391 Seconds
(without alignments)
10346.286 Million cell updates/sec

Title: US-10-607-806-1-T9182_COPY_8700_9600
Perfect score: 901
Sequence: 1 tggctcataattataaaat.....gagctcggcaggagatc 901

Scoring table: IDENTITY NUC
Gapop 10_0 , Gapext 1.0

Searched: 4996997 seqs, 332346308 residues 9993994
Total number of hits satisfying chosen parameters:

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

- Database : N_Geneseq_21.*
- 1: Geneseqn1980s.*
 - 2: Geneseqn1990s.*
 - 3: Geneseqn2000s.*
 - 4: Geneseqn2001s.*
 - 5: Geneseqn2001bs.*
 - 6: Geneseqn2002s.*
 - 7: Geneseqn2002bs.*
 - 8: Geneseqn2003s.*
 - 9: Geneseqn2003bs.*
 - 10: Geneseqn2003cs.*
 - 11: Geneseqn2003ds.*
 - 12: Geneseqn2004s.*
 - 13: Geneseqn2004bs.*
 - 14: Geneseqn2005s.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	901	100.0	12174	12	ADJ09983 Human pho
2	900.6	100.0	12174	12	Adi35082 Human PLA
3	899.4	99.8	13612	6	Abk47376 Human Ph
c	555.4	61.6	42360	13	ABD33466 Human can
5	313.6	34.8	22976	2	AX83426 Genomic r
6	313.6	34.8	23187	3	AA50273 Human lip
7	313.6	34.8	23187	4	AAF62331 Human lep
8	313.6	34.8	41936	6	AB167924 Ovary can
9	313.6	34.8	56580	13	ADT25932 Wild type
10	313.6	34.8	56580	13	ADU23759 Human ATP
11	312.6	34.7	21721	2	AX83427 Human lip
c	312.4	34.7	37940	13	ADS19371 Human PIP
c	312.4	34.7	38000	12	ADJ26093 Human PIP
c	310	34.4	28818	4	AA135901 Human mus
c	310	34.4	28818	8	ABX58889 cDNA enco
c	310	34.4	28818	12	ADJ29639 Human mus
17	308.2	34.2	19686	11	ACN44170 Human gen
18	307.2	34.1	69770	10	ADC86870 Human GPC
c	307	34.1	234882	14	ADZ13715 Human can

c	20	307	34.1	246940	12	ADQ59422	Human can
c	21	305.2	33.9	23434	8	ACC00503	Human his
c	22	304	33.7	260160	12	ADQ20017	Human sof
c	23	303.8	33.7	32250	4	AA105627	Human rep
c	24	303.2	33.7	114693	8	AA48308	Human tra
c	25	302.8	33.6	14598	4	AAK80690	Human imm
c	26	302.8	33.6	14598	4	AAK79627	Human imm
c	27	302.6	33.6	174472	14	ADZ13139	Human can
c	28	302.6	33.6	174703	11	ACN44738	Human gen
c	29	302.2	33.5	38918	10	ADC87242	Human GPC
c	30	301.8	33.5	2611	6	ABS76421	cDNA enco
c	31	301.8	33.5	2611	14	ADZ51348	Nucleotid
c	32	301.6	33.5	59588	10	ADC87426	Human GPC
c	33	301.6	33.5	301477	13	ABD33362	Human can
c	34	300.4	33.3	167163	10	ADE82948	Human pvt
c	35	300.2	33.3	37160	6	ABK84372	Human cdn
c	36	299.4	33.2	1341	6	ABK14184	Human spl
c	37	297.4	33.0	110000	11	ACN44014	Continuation (3 of
c	38	297.2	33.0	16869	3	ABN97976	Human ret
c	39	297.2	33.0	56093	6	ABL61744	Colon ade
c	40	297.2	33.0	119501	12	ADI29095	Human MAR
c	41	297	33.0	133632	11	ACN45054	Human gen
c	42	297	33.0	133642	14	ADZ13285	Human can
c	43	296.4	32.9	110000	14	AEA61163	Continuation (2 of
c	44	295.8	32.8	103747	6	ABQ88139	Human ost
c	45	295.2	32.8	8458	4	AAK72176	Human imm

ALIGNMENTS

RESULT 1

ADJ09983

ID ADJ09983 standard; DNA; 12174 BP.

XX AC ADJ09983;

XX DT 17-JUN-2004 (first entry)

XX DE Human phospholipase A2 (PLA2G1B) DNA SeqID 1.

XX KW human; gene; ds; fat reduction; fat deposition; phospholipase A2; PLA2G1B; chromosome 12q24; single nucleotide polymorphism; SNP; appetite suppressant; lipase inhibitor; exercise regimen; obesity; non-insulin dependent diabetes mellitus; NIDDM; cardiovascular disorder; hypertension; antidiabetic.

XX OS Homo sapiens.

XX	Key	Location/Qualifiers
FT	variation	replace(436,c)
FT	variation	/*tag= a
FT	variation	/standard_name= "Single nucleotide polymorphisms"
FT	variation	replace(839,a)
FT	variation	/*tag= b
FT	variation	/standard_name= "Single nucleotide polymorphisms"
FT	variation	replace(4050,a)
FT	variation	/*tag= c
FT	variation	/standard_name= "Single nucleotide polymorphisms"
FT	variation	replace(4689,t)
FT	variation	/*tag= d
FT	variation	/standard_name= "Single nucleotide polymorphisms"
FT	variation	replace(6282,a)
FT	variation	/*tag= e
FT	variation	/standard_name= "Single nucleotide polymorphisms"
FT	variation	replace(6358,c)
FT	variation	/*tag= f
FT	variation	/standard_name= "Single nucleotide polymorphisms"
FT	variation	replace(6653,t)
FT	variation	/*tag= g
FT	variation	/standard_name= "Single nucleotide polymorphisms"
FT	variation	replace(7256,t)
FT	variation	/*tag= h

FT /standard_name= "Single nucleotide polymorphisms"
 FT replace(7300,a)
 FT /*tag= i
 FT /standard_name= "Single nucleotide polymorphisms"
 FT replace(7301,c)
 FT /*tag= j
 FT /standard_name= "Single nucleotide polymorphisms"
 FT replace(7328,g)
 FT /*tag= k
 FT /standard_name= "Single nucleotide polymorphisms"
 FT replace(8062,c)
 FT /*tag= l
 FT /standard_name= "Single nucleotide polymorphisms"
 FT replace(9182,t)
 FT /*tag= m
 FT /standard_name= "Single nucleotide polymorphisms"
 FT replace(11649,c)
 FT /*tag= n
 FT /standard_name= "Single nucleotide polymorphisms"
 XX WO2004002296-A2.
 XX
 XX 08-JAN-2004.
 XX
 XX 27-JUN-2003; 2003WO-US020831.
 XX
 XX 27-JUN-2002; 2002US-0392362P.
 XX (SEQU-) SEQUENOM INC.
 XX
 XX Adam GIR, Langdown ML, Denissenko MF, Dennis E, Cantor C;
 XX Rubin B;
 XX
 XX WPI; 2004-071944/07.
 XX P-PSDB; ADJ09984.
 XX
 XX Identifying a candidate therapeutic for fat reduction, useful for
 XX treating diabetes, by introducing a test molecule to a system comprising
 XX PLA2G1B protein or nucleic acid, and determining the presence of
 XX interaction between the compounds.
 XX
 XX Claim 1; SEQ ID NO 1; 116pp; English.
 XX
 XX This invention relates to a novel candidate therapeutic agent useful for
 XX fat reduction and disorders related to fat depositions. Specifically, it
 XX refers to polymorphic variations in the phospholipase A2 (PLA2G1B) DNA,
 XX which is located on chromosome 12q24 and has been associated with central
 XX fat deposition. The present invention describes methods to detect the
 XX presence or absence of these single nucleotide polymorphisms of PLA2G1B,
 XX in particular G7328A and T9182G, and subsequently provide treatment that
 XX reduces fat deposition. This treatment may consist of an appetite
 XX suppressant, a lipase inhibitor, a phospholipase inhibitor, an exercise
 XX regimen, a dietary regimen, psychological counselling, psychotherapy or a
 XX psychotherapeutic. Accordingly, PLA2G1B is a target for reducing fat
 XX deposition and it can be used to treat both obesity and non-insulin
 XX dependent diabetes mellitus (NIDDM), as well as cardiovascular disorders
 XX such as hypertension. As such, it exhibits antidiabetic activity. This
 XX polynucleotide sequence is the human PLA2G1B DNA of the invention.
 XX
 XX Sequence 12174 BP; 3220 A; 2996 C; 2739 G; 3219 T; 0 U; 0 Other;
 XX
 XX Query Match 100.0%; Score 901; DB 12; Length 12174;
 XX Best Local Similarity 100.0%; Pred. No. 3e-187;
 XX Matches 901; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 XX
 XX 1 TGTTCATTAATAAATAAATCACTCAGTTCGGCACAGTGACTCAAGCTGTAAACACAGT 60
 XX ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
 XX 8700 TGTTCATTAATAAATAAATAAATCACTCAGTTCGGCACAGTGACTCAAGCTGTAAACACAGT 8759
 XX ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
 XX 61 ACTTTGGAAGTCCCAAGTGGGTGGATCATTCTGAGGTGAGAAGTTCGAGACCCAGCTGGTC 120
 XX ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
 XX 8760 ACTTTGGAAGTCCCAAGTGGGTGGATCATTCTGAGGTGAGAAGTTCGAGACCCAGCTGGTC 8819
 XX ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||

Qy 121 AATATGCTGAAGCCCTATCTCTACTAAAAATACAAAAATAGCTGGGTAGTGATGCAT 180
 ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
 Db 8820 AATATGCTGAAGCCCTATCTCTACTAAAAATACAAAAATAGCTGGGTAGTGATGCAT 8879
 ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
 Qy 181 GCCTGTAGTCCCAGCTACTCGGAGGCTGAGGCAAGAGAATTCCTTGAACCTGGGAGGCA 240
 ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
 Db 8880 GCCTGTAGTCCCAGCTACTCGGAGGCTGAGGCAAGAGAATTCCTTGAACCTGGGAGGCA 8939
 ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
 Qy 241 GAGGTTTCAGTGCAGCCAGATCCACCACTGCACTCCAGCCTGGGCCGACACAGGAGACT 300
 ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
 Db 8940 GAGGTTTCAGTGCAGCCAGATCCACCACTGCACTCCAGCCTGGGCCGACACAGGAGACT 8999
 ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
 Qy 301 CTATCTCAAAAAATAAATAAATAAAGATCGGAGAGAAACAAAACTAATAAGAT 360
 ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
 Db 9000 CTATCTCAAAAAATAAATAAATAAAGATCGGAGAGAAACAAAACTAATAAGAT 9059
 ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
 Qy 361 TCCTGAAGGTAAGCAGAGATACGTAATAATATATGTAATAAGTTTAAATGCAATTTAACT 420
 ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
 Db 9060 TCCTGAAGGTAAGCAGAGATACGTAATAATATATGTAATAAGTTTAAATGCAATTTAACT 9119
 ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
 Qy 421 GTAATCTATTGTTTATTTTGGTTTAAAGATTAACCAAGCCAAAGTAATCAACTTCAA 480
 ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
 Db 9120 GTAATCTATTGTTTATTTTGGTTTAAAGATTAACCAAGCCAAAGTAATCAACTTCAA 9179
 ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
 Qy 481 ACTCTACATAAATATCTATTATGGAAGTGGAGGCACTCTAATAATCCTACTACCAAGA 540
 ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
 Db 9180 ACTCTACATAAATATCTATTATGGAAGTGGAGGCACTCTAATAATCCTACTACCAAGA 9239
 ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
 Qy 541 TAACAGGTTACATATTCCTCCAGATTTTGGGGGATACACTAGCTTTTTTATTGGGAA 600
 ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
 Db 9240 TAACAGGTTACATATTCCTCCAGATTTTGGGGGATACACTAGCTTTTTTATTGGGAA 9299
 ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
 Qy 601 AATTTCATGTCAGGCATACCTAATTTTCTAAATGCTATGTAGTATTCATTAAAGG 660
 ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
 Db 9300 AATTTCATGTCAGGCATACCTAATTTTCTAAATGCTATGTAGTATTCATTAAAGG 9359
 ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
 Qy 661 ATGTTCCATAATTTTAAATAACATGCTTTAAAGTAGAGAAACTAGTTGGGCATGGTG 720
 ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
 Db 9360 ATGTTCCATAATTTTAAATAACATGCTTTAAAGTAGAGAAACTAGTTGGGCATGGTG 9419
 ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
 Qy 721 CTCACGCTGTATCCCAGCACTTTGGGAGGCGGCAAAATGGATCATTGAGGTCCGGA 780
 ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
 Db 9420 CTCACGCTGTATCCCAGCACTTTGGGAGGCGGCAAAATGGATCATTGAGGTCCGGA 9479
 ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
 Qy 781 GTTTCAGACCCAGCTGGACACATGATCAACACCCCTCTCTAATAAAAAATACAAAAATTA 840
 ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
 Db 9480 GTTTCAGACCCAGCTGGACACATGATCAACACCCCTCTCTAATAAAAAATACAAAAATTA 9539
 ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
 Qy 841 GCTGGGCATGGTGGCAAGCACCTGTAGTCCCAGCTACTCAGGAGTCTGAGGCAGGAGTAT 900
 ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
 Db 9540 GCTGGGCATGGTGGCAAGCACCTGTAGTCCCAGCTACTCAGGAGTCTGAGGCAGGAGTAT 9599
 ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
 Qy 901 C 901
 |||||
 Db 9600 C 9600
 |||||
 XX RESULT 2
 XX ADI35082
 XX ID ADI35082 standard; DNA; 12174 BP.
 XX
 XX AC ADI35082;
 XX
 XX XX 22-APR-2004 (first entry)
 XX
 XX DE Human PLA2G1B nucleotide sequence.
 XX
 XX XX
 XX KW PLA2G1B ; fat deposition; leanness; polymorphism;
 XX KW non-insulin dependent diabetes mellitus; NIDDM; hyperinsulinemia;
 XX KW hypertension; glucose intolerance; dyslipidemia; hypercoagulability;
 XX KW microalbuminuria; human; gene; ds.
 XX
 XX OS Homo sapiens.

RESULT 4
ABD33466/c
ID ABD33466 standard; DNA; 42360 BP.
XX
AC ABD33466;
XX
DT 18-NOV-2004 (first entry)
XX
DE Human cancer-associated (CA) gene HD07-089.
XX
KW Human; cancer-associated protein; CAP; cancer-associated gene; CA; gene;
KW ds; cancer; cytostatic.
XX
OS Homo sapiens.
XX
PN WO2004058146-A2.
XX
PD 15-JUL-2004.
XX
PF 15-DEC-2003; 2003WO-US040081.
XX
PR 17-DEC-2002; 2002US-00322281.
XX
PA (SAGR-) SAGRES DISCOVERY INC.
XX
PI Morris DW, Malandro MS;
XX
DR WPI; 2004-499109/47.
XX
PT Novel human cancer associated protein encoded within open reading frame
PT of cancer associated gene, useful as targets for diagnosing cancer.
XX
PS Claim 16; SEQ ID NO 618; 182pp; English.
XX
CC The invention relates to cancer-associated proteins (CAP) and the cancer-
CC associated (CA) nucleic acids encoding them. The invention also relates
CC to a method for treating cancers involving administering to a patient an
CC inhibitor of CAP, and a method of screening for anticancer activity in a
CC potential drug involving providing a cell that expresses a CA gene,
CC contacting a tissue sample derived from a cancer cell with an anticancer
CC drug candidate and monitoring the effect of the anticancer drug candidate
CC on expression of the CA gene. The CAP proteins are useful for detecting
CC cancer associated with expression of a CAP protein in a test cell sample
CC and for screening for a bioactive agent capable of modulating the
CC activity of a CAP protein. The CA nucleic acids are useful for diagnosing
CC cancer, involving determining the expression of a CA nucleic acid in a
CC tissue. This sequence represents a human CA gene of the invention. Note:
CC The sequence data for this patent did not form part of the printed
CC specification, but was obtained in electronic format directly from WIPO
CC at ftp.wipo.int/pub/published_pct_sequences
SQ Sequence 42360 BP; 11698 A; 9658 C; 10445 G; 10559 T; 0 U; 0 Other;

Query Match 61.6%; Score 555.4; DB 13; Length 42360;
Best Local Similarity 99.8%; Pred. NO. 1.6e-111;
Matches 556; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 345 ACAAACTAATAAGATTCTCTGAAGTAAAGCAGATACGTAATATATATGTAATAAGTT 404
DB 42360 ACAAACTAATAAGATTCTCTGAAGTAAAGCAGATACGTAATATATATGTAATAAGTT 42301

QY 405 TAAATGCAATTTTAACTGTAATCTTATGTTTATTTGTTTATAAAGTAAACAGCCAAA 464
DB 42300 TAAATGCAATTTTAACTGTAATCTTATGTTTATTTGTTTATAAAGTAAACAGCCAAA 42241

QY 465 AGTAATGCACTTCAAACTCTACATAAATATCTATTATGGAAGTGAAGSCATCTATAA 524
DB 42240 AGTAATGCACTTCAAACTCTACATAAATATCTATTATGGAAGTGAAGSCATCTATAA 42181

QY 525 TCCTACTACCCAAAGATAACAGTTACATATTCCTCCAGATTTTGGGGCATACTAGC 584
|||||

DB 42180 TCCTACTACCCAAAGATAACAGTTACATATTCCTCCAGATTTTGGGGCATACTAGC 42121
QY 585 TTTTATTTATTTGGGAAAATTTCCATGTCAGGCATACCTAATTTTCTTAATGTCATGT 644
DB 42120 TTTTATTTATTTGGGAAAATTTCCATGTCAGGCATACCTAATTTTCTTAATGTCATGT 42061
QY 645 AGTATTCATTTAAGGATGTTTCCATAATTTTAAAAATACATGCTTTTAAAGTAGAGAACT 704
DB 42060 AGTATTCATTTAAGGATGTTTCCATAATTTTAAAAATACATGCTTTTAAAGTAGAGAACT 42001
QY 705 AGTTGGGCATGTTGGCTCAGGCTGTATCCAGACATTTTGGAGGCGGAGGCAAAATGGA 764
DB 42000 AGTTGGGCATGTTGGCTCAGGCTGTATCCAGCACATTTTGGAGGCGGAGGCAAAATGGA 41941
QY 765 TCACCTTGAGGTCCGGAGTTTGAGACACAGCCTGGGCAACATGATGAACACACCCCTCTTAAT 824
DB 41940 TCACCTTGAGGTCCGGAGTTTGAGACACAGCCTGGGCAACATGATGAACACACCCCTCTTAAT 41881
QY 825 AAAAATACAAAAATTTAGCTGGGCATGTTGGCAAGCACCTGTAGTCCAGCTACTCAGGAG 884
DB 41880 AAAAATACAAAAATTTAGCTGGGCATGTTGGCAAGCACCTGTAGTCCAGCTACTCAGGAG 41821
QY 885 TCTGAGGCGAGGTATC 901
DB 41820 TCTGAGGCGAGGTATC 41804

RESULT 5
AAX83426
ID AAX83426 standard; DNA; 22976 BP.
XX
AC AAX83426;
XX
DT 31-AUG-1999 (first entry)
XX
DE Genomic region containing human lipolysis stimulated receptor gene.
XX
KW Human; lipolysis stimulated receptor; LSR; lipoprotein; free fatty acid;
KW cytokine; probe; primer; amplification; hybridisation; detection; leptin;
KW allele; variant; mutation; deletion; loss of heterozygosity; chylomicron;
KW transgenic animal; gene expression; triglyceride; eating disorder;
KW obesity; atheromatosis; atherosclerosis; hypertension; diabetes;
KW anorexia; metabolism; ds.
XX
OS Homo sapiens.
XX
PN WO9907737-A2.
XX
PD 18-FEB-1999.
XX
PF 06-AUG-1998; 98WO-IB001257.
XX
PR 06-AUG-1997; 97FR-00010088.
PR 22-APR-1998; 98FR-00005032.
XX
XX (GEST) GENSET.
XX (INRM) INSERM INST NAT SANTE & RECH MEDICALE.
XX
PI Bihain B, Bougueleret L, Yen-Potin F;
XX
DR WPI; 1999-190035/16.
XX
PT Lipolysis stimulated receptor involved in leptin metabolism - and
PT controlling its activity for treatment of obesity, anorexia or abnormal
PT cytokine metabolism.
XX
PS Claim 40; Page 222-242; 279pp; French.
XX
CC This sequence represents the genomic region containing the coding region
CC for the human lipolysis stimulated receptor (LSR). The novel LSR binds
CC lipoproteins in presence of free fatty acids (FFA) and binds cytokines in
CC absence of FFA. The nucleic acid and its fragments are used as probes and
CC primers for detection and/or amplification of LSR genes; for production

CC of recombinant LSR; for detecting allelic variants, mutations, deletions,
CC loss of heterozygosity and genetic abnormalities in the gene. LSR,
CC recombinant cells and transgenic animals are used to screen for chemical
CC interacting with LSR, also to study expression and activity of LSR and
CC its interactions. The chemicals, and leptin, are used to modulate the
CC number of LSR in a cell, its recycling rate and/or specificity of
CC receptor activity, particularly for reducing the level of leptin,
CC lipoproteins, chylomicrons and/or triglycerides. The chemicals are thus
CC useful for treating eating disorders, particularly obesity (and related
CC diseases such as atheromatosis, atherosclerosis, hypertension and
CC diabetes) or anorexia, also disease associated with abnormal cytokine
CC metabolism
XX
SQ Sequence 22976 BP; 5120 A; 6191 C; 6677 G; 4988 T; 0 U; 0 Other;
Query Match 34.8%; Score 313.6; DB 2; Length 22976;
Best Local Similarity 62.4%; Pred. No. 1.2e-58;
Matches 545; Conservative 0; Mismatches 319; Indels 10; Gaps 3;
QY 29 TGGGCACAGTGACTCAAGCCTGTAAACACAGTACTTTGGAAGTCAAGGTGGGTGGATCA 88
DB 5491 TGGGCACGCTGGCTCACGCTCTGTAATCTCAGCACTTTGGGAGGCCAAGTGGGTGGATCA 5550
QY 89 CTTGAGGTGAGAAAGTTCGAGACCAGCCTGTCATATGTTGTAACCTATCTCTACTAAA 148
DB 5551 CCTGAGGTGAGGATTCAGACACAGCCTGGCCAAACATGGTGAACCCCGTCTCTACTAAA 5610
QY 149 AATACAAAATTAAGTCTGGGTGAGTGCATGCCCTGTAGTCCCAAGCTACTCGGGAGGCT 208
DB 5611 AATACAAAGATTAGCCAGGCGATGGTGGCACATGCCCTGTAAATCCCAAGCTACTCAGGAGGCT 5670
QY 209 GAGGCAAGAGAAATGCTTGAACCTGGGAGGACAGAGTTCGAGTACGAGAGTCCACCA 268
DB 5671 GAGGCATGAGATCGCTTGAACCCAGGAGGACAGAGTTCGAGTACGAGTGCACCA 5730
QY 269 CTGCACTCAGCCTGGGCGACACAGGAGACTCTATCTCAAAAATAAATAAATAAAT 328
DB 5731 CTGTACTGCACTCTGGGCGACAGAGTGCATCTGTCTCAAAAATAAATAAATAAATAA 5790
QY 329 AAAGGATCGAGAGAAACAAACTAATAAGATTCTCTGAAGGTGAAGCAGAGATACGTAAT 388
DB 5791 GACTCGTCAAGGTATAAGATGTCAGAGAGTACTAAGTGTGCAAG-----AAAA 5842
QY 389 TATATGTAATAAGTTTAATGCAATTTAACTGTGTAATCTTATTTATTTGGTTATAA 448
DB 5843 TAACACCAAGCTGGGTGCATTTGGCTCATGCTGTAAATTTAGCACTTTGGGAGGCCAAG 5902
QY 449 AAGTAAACAGCCAAAGTAATGCAACTTCAACTCTACATAATATCTATTATGGAAG 508
DB 5903 GCAGGAGGATCACTTGAGCCTAGGAGTTTGAGACCAGCCTGGACCAACAAATGAGACCCC 5962
QY 509 TGAAGGCATCTATATCTCTACTTACCAAGATAACCAAGTATACATATCTCCAGATTTT 568
DB 5963 ATGCTACAAAATTTTAAATTTTAAATTTTAAATTTAGCTGGGCATGGTGGCATGGCTGG 6022
QY 569 TGGGGCATACACTAGCTTTTTTTTATTGGGAAATTTCCATGTGCAAGCATACCTAATTT 628
DB 6023 TCCCGGCTGCTCAGGAGGCTGAGGTGGGAGGATTGCTTGGGCTTGAGAGGTCAAGGCTTC 6082
QY 629 TTCTAAATGCTATGATGATTTCCATTAAGGATGTTCCATATTTTAAATATACATGCT 688
DB 6083 AGTGAGTCAATGATCGTGCCACTGCTTCCAGCCTGGGTGA-CAGAGTGAGACCTGCTCTT 6141
QY 689 TTAAGTAGAGAACTAGGTTGGGCATGGTGGCTCACGCTGT-ATCCCAGCACTTTGGG 747
DB 6142 GAAATGAAAGAAAATAGGCTGGCGCAGTGGCTCACCTGTGTAATCCCAGCACTTTGGG 6201
QY 748 AGGCCGAGGCAAAATGGATCACTTGAAGTCCGAGTTTGAGACCAGCCTGGAACAATGAT 807
DB 6202 AGGCCGAGGTGGGTGGATCACTTCACTGAGTCAAGGATCGAGACCAGCCTGGCCAACTGGT 6261
QY 808 GAAACACCTCTCTAATAAATAAATAAATAGCTGGGCATGGTGGCAGCAGCCTGTAG 867

DB 6262 GAAATCCCATCTCTACTAAAAATACAAAATTTAGCCGGCGGTGGTGGCGCCTGTAA 6321
QY 868 TCCAGCTACTCAGGAGTCTTGAGGAGGAGTATC 901
DB 6322 TCCAGCTACTCAGGAGGCTGAGGAGGAGATC 6355
RESULT 6
AAAA50273
ID AAA50273 standard; DNA; 23187 BP.
XX AC
XX AAA50273;
XX
DT 07-NOV-2000 (first entry)
XX Human lipolysis stimulated receptor (LSR) gene.
DE DE
XX
XX Lipolysis stimulated receptor; LSR; chromosome 19q13.1; human;
KW single nucleotide polymorphism; biallelic marker; obesity;
KW atherosclerosis; insulin resistance; hypertension; hyperlipidemia;
KW hypertriglyceridemia; cardiovascular disease; microangiopathy;
KW syndrome X; diagnosis; therapy; genotyping; ds.
XX
OS Homo sapiens.
XX
XX
FH Key
FT 5'UTR
FT Location/Qualifiers
FT 1..2000
FT /*tag= a
FT /note= "potential 5' regulatory region"
FT complement(523..544)
FT /*tag= bj
FT primer_bind
FT /note= "upstream amplification primer 17-2"
FT replace(606,T)
FT /*tag= am
FT /standard_name= "single nucleotide polymorphism"
FT /note= "marker A.1"
FT replace(818,G)
FT /*tag= u
FT /standard_name= "single nucleotide polymorphism"
FT /note= "marker 17-2-297 (A31)"
FT complement(946..963)
FT /*tag= bl
FT /note= "upstream amplification primer 99-4576"
FT 1047..1068
FT /*tag= bk
FT /note= "downstream amplification primer 17-2"
FT complement(1096..1115)
FT /*tag= bn
FT /note= "upstream amplification primer 9-19"
FT replace(1243,T)
FT /*tag= v
FT /frequency= "0.15"
FT /standard_name= "single nucleotide polymorphism"
FT /note= "marker 9-19-148 (A10)"
FT replace(1374,A)
FT /*tag= w
FT /standard_name= "single nucleotide polymorphism"
FT /note= "marker 9-19-256 (A32)"
FT 1385..1402
FT /*tag= bm
FT /note= "downstream amplification primer 99-4576"
FT replace(1401,T)
FT /*tag= x
FT /frequency= "0.12"
FT /standard_name= "single nucleotide polymorphism"
FT /note= "marker 9-19-307 (A11)"
FT replace(1534..1535,C)
FT /*tag= y
FT /standard_name= "single nucleotide polymorphism"
FT /note= "marker 9-19-442 (A12)"
FT complement(1602..1621)
FT /*tag= bp
FT /note= "upstream amplification primer 9-20"
FT


```
FT primer_bind 1616.1635
FT /tag= bo
FT /note= "downstream amplification primer 9-19"
FT replace(1788,A)
FT /tag= z
FT /standard_name= "single nucleotide polymorphism"
FT /note= "marker 9-20-187 (A13)"
FT 2001.2356
FT /tag= b
FT /number= 1
FT complement(2036..2053)
FT /tag= br
FT /note= "upstream amplification primer 99-4557"
FT 2062..2081
FT /tag= bw
FT /note= "downstream amplification primer 9-21"
FT 2074..2093
FT /tag= bq
FT /note= "downstream amplification primer 9-20"
FT complement(2084..2102)
FT /tag= bt
FT /note= "upstream amplification primer 9-1"
FT 2104..21082
FT /tag= bi
FT /note= "Contains introns, encodes AAY95822, AAY95823 and
FT AAY95824 as a result of alternative splicing"
FT 2357..3539
FT /tag= c
FT /number= 1
FT replace(2391,G)
FT /tag= aa
FT /frequency= "0.24"
FT /standard_name= "single nucleotide polymorphism"
FT /note= "marker 9-1-308 (A14)"
FT complement(2470..2489)
FT /tag= bv
FT /note= "upstream amplification primer 9-21"
FT 2483..2500
FT /tag= bu
FT /note= "downstream amplification primer 9-1"
FT 2563..2580
FT /tag= bs
FT /note= "downstream amplification primer 99-4557"
FT complement(3455..3474)
FT /tag= bx
FT /note= "upstream amplification primer 9-3"
FT 3540..3884
FT /tag= d
FT /number= 2
FT complement(3775..3792)
FT /tag= bz
FT /note= "upstream amplification primer 99-4558"
FT replace(3778,T)
FT /frequency= "0.29"
FT /standard_name= "single nucleotide polymorphism"
FT /note= "marker 9-3-324 (A15)"
FT 3882..3901
FT /tag= by
FT /note= "downstream amplification primer 9-3"
FT 3885..12162
FT /tag= e
FT /number= 2
FT 4336..4356
FT /tag= ca
FT /note= "downstream amplification primer 99-4558"
FT 4444..44638
FT /tag= cd
FT /note= "downstream amplification primer 99-14419"
FT replace(4498,T)
FT /tag= ac
FT /frequency= "0.22"
FT /standard_name= "single nucleotide polymorphism"
```

```
FT primer_bind /note= "marker 99-14419-424 (A16)"
FT complement(4902..4920)
FT /tag= cb
FT /note= "upstream amplification primer 99-14419"
FT replace(5141,G)
FT /tag= an
FT /standard_name= "single nucleotide polymorphism"
FT /note= "marker A'2"
FT complement(6638..6655)
FT /tag= ce
FT /note= "upstream amplification primer 99-4577"
FT 7072..7089
FT /tag= cf
FT /note= "downstream amplification primer 99-4577"
FT replace(7428,C)
FT /tag= ao
FT /standard_name= "single nucleotide polymorphism"
FT /note= "marker A'3"
FT complement(7995..8012)
FT /tag= cg
FT /note= "upstream amplification primer 99-4559"
FT replace(8394,G)
FT /tag= ap
FT /standard_name= "single nucleotide polymorphism"
FT /note= "marker A'4"
FT 8576..8593
FT /tag= ch
FT /note= "downstream amplification primer 99-4559"
FT replace(8704,C)
FT /tag= aq
FT /standard_name= "single nucleotide polymorphism"
FT /note= "marker A'5"
FT replace(9028,A)
FT /tag= ar
FT /standard_name= "single nucleotide polymorphism"
FT /note= "marker A'6"
FT complement(9622..9639)
FT /tag= ci
FT /note= "upstream amplification primer 99-3148"
FT replace(9949..9958,TA)
FT /tag= ab
FT /note= "marker A'7"
FT complement(9964..9981)
FT /tag= ck
FT /note= "upstream amplification primer 99-4560"
FT replace(9977,C)
FT /tag= at
FT /standard_name= "single nucleotide polymorphism"
FT /note= "marker A'8"
FT replace(10021,A)
FT /tag= au
FT /standard_name= "single nucleotide polymorphism"
FT /note= "marker A'9"
FT 10023..10040
FT /tag= cj
FT /note= "downstream amplification primer 99-3148"
FT 10492..10512
FT /tag= cn
FT /note= "downstream amplification primer 99-14411"
FT 10546..10563
FT /tag= cl
FT /note= "downstream amplification primer 99-4560"
FT complement(10996..11015)
```

```
Query Match 34.8%; Score 313.6; DB 3; Length 23187;
Best Local Similarity 62.4%; Pred. No. 1.2e-58;
Matches 545; Conservative 0; Mismatches 319; Indels 10; Gaps 3;

Qy 29 TGGGCACAGTACTCAAGCCTGTAAACACAGTACTTTGGAGTCCAGGTGGGTGATCA 88
| | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 5595 TGGGCACGGTGGCTCAGCTCTGAATCTCAGCACTTTGGGAGGCCAAGGTGGGTGATCA 5654
| | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |

Qy 89 CTTGAGGTGAGAGTTTCGAGACCAGCCTGGTCAATATGGTGAACCCCTATCTCTACTAAA 148
```


Db 6007 GCAGAGGATCACTTGAGCGCTAGGAGTTTGACACAGCCTGGACACAAAATGAGACCCC 6066
QY 509 TGGAGGCGATCTATATCTACTACCCAAAAGATAACCAAGTTTACATATTTCTCCAGATTTT 568
Db 6067 ATGCTACAAAAATTTTAAAAATTTTAAAAATTTAGCTGGGCATGGTGCCTGTGG 6126
QY 569 TGGGCGATACACTAGCTTTTATTTTGGGAAAATTTCCATGTGCGAGCATACCTAATTT 628
Db 6127 TCCCGGCTGCTCAGGAGGCTGAGGTGGGAGGATTTGCTTGGGCTTGAGAGGTCAGAGGCTTC 6186
QY 629 TTCTAAATGCTCTATGATGATTTCCATTTAAGGATGTTCCATAATTTTAAAAATACATGCT 688
Db 6187 AGTGAGTCATGATCGTGCACCTGCAATTCAGCCTGGGTGA-CAGAGTGAGACCCCTGTCTT 6245
QY 689 TTAAGTGTAGAAAACTAGGTTGGGCATGGTGGCTCACGCCTGT-ATCCAGCAGCTTTGGG 747
Db 6246 GAAATGAAAAGAAAATAGCTGGCGCAGTGGCTCACACCTGTATATCCAGCAGCTTTGGG 6305
QY 748 AGGCGGAGGCAATGGATCACTTGAGGTCCGGAGTTTGAGACCAGCCTGGACAAATGAT 807
Db 6306 AGGCGGAGGTGGGTGGATCACCTGAGGTGAGGATCGAGATCGAGATCGGCAACATGGT 6365
QY 808 GAACACACCTCTCTAATAAAAAATACAAAAATTAGCTGGGCATGGTGGCAAGCACCTGTAG 867
Db 6366 GAAATCCCATCTCTACTAAAAATACAAAAATTTAGCGCGGCTGGTGGCGGCTGTAA 6425
QY 868 TCCAGCTACTCAGGAGTCTGAGGCGAGGATATC 901
Db 6426 TCCAGCTACTCGGAGGCTGAGGCGAGGATC 6459
XX ABL67924;
AC ABL67924;
AT 15-MAY-2002 (first entry)
DT Ovary cancer related gene sequence SEQ ID NO:6261.
DE Human; cancer; colon; breast; ovary; oesophagus; kidney; thyroid;
KW stomach; lung; prostate; pancreas; carcinoma; antitumour; cancerous;
KW cytostatic; gene therapy; antineoplastic; Wilm's tumour; adenocarcinoma;
KW gene; ds.
XX Homo sapiens.
XX WO200194629-A2.
XX 13-DEC-2001.
XX 30-MAY-2001; 2001WO-US010838.
XX 05-JUN-2000; 2000US-0209473P.
PR 05-JUN-2000; 2000US-0209531P.
PR 18-SEP-2000; 2000US-0233133P.
PR 18-SEP-2000; 2000US-0233161P.
PR 20-SEP-2000; 2000US-0234009P.
PR 20-SEP-2000; 2000US-0234034P.
PR 20-SEP-2000; 2000US-0234052P.
PR 22-SEP-2000; 2000US-0234509P.
PR 22-SEP-2000; 2000US-0234567P.
PR 25-SEP-2000; 2000US-0234923P.
PR 25-SEP-2000; 2000US-0235077P.
PR 25-SEP-2000; 2000US-0235082P.
PR 25-SEP-2000; 2000US-0235134P.
PR 25-SEP-2000; 2000US-0235280P.
PR 26-SEP-2000; 2000US-0235637P.
PR 26-SEP-2000; 2000US-0235638P.
PR 27-SEP-2000; 2000US-0235711P.

PR 27-SEP-2000; 2000US-0235720P.
PR 27-SEP-2000; 2000US-0235840P.
PR 27-SEP-2000; 2000US-0235863P.
PR 28-SEP-2000; 2000US-0236028P.
PR 28-SEP-2000; 2000US-0236032P.
PR 28-SEP-2000; 2000US-0236033P.
PR 28-SEP-2000; 2000US-0236034P.
PR 28-SEP-2000; 2000US-0236109P.
PR 28-SEP-2000; 2000US-0236111P.
PR 29-SEP-2000; 2000US-0236842P.
PR 29-SEP-2000; 2000US-0236891P.
PR 02-OCT-2000; 2000US-0237172P.
PR 02-OCT-2000; 2000US-0237173P.
PR 02-OCT-2000; 2000US-0237278P.
PR 02-OCT-2000; 2000US-0237294P.
PR 02-OCT-2000; 2000US-0237295P.
PR 02-OCT-2000; 2000US-0237316P.
PR 03-OCT-2000; 2000US-0237425P.
PR 03-OCT-2000; 2000US-0237598P.
PR 03-OCT-2000; 2000US-0237604P.
PR 03-OCT-2000; 2000US-0237606P.
PR 03-OCT-2000; 2000US-0237608P.
PR 01-NOV-2000; 2000US-0244867P.
PR 01-NOV-2000; 2000US-0245084P.
XX (AVAL-) AVALON PHARM.
XX Young PE, Augustus M, Carter KC, Ebner R, Endress G, Horrigan S;
PI Soppet DR, Weaver Z;
XX WPI; 2002-188264/24.
XX Screening for anti-neoplastic agent involves exposing cells to a chemical
PT agent to be tested for anti-neoplastic activity, and determining a change
PT in expression of a gene of a signature gene set.
XX Claim 1; SEQ ID NO 6261; 44pp; English.
XX The present invention describes a method (M1) for screening for an anti-
CC neoplastic agent. The method involves exposing cells to a chemical agent
CC to be tested for anti-neoplastic activity, determining a change in
CC expression of at least one gene (I) of a signature gene set, where (I)
CC comprises a sequence (S) selected from 8447 sequences (given in ABL61664
CC to ABL70110), or is at least 95% identical to (S), where a change in
CC expression is indicative of anti-neoplastic activity. (I) has cytostatic
CC activity and can be used in gene therapy. M1 can be used for screening an
CC anti-neoplastic agent, and can be used for producing a product which is
CC the data collected with respect to the anti-neoplastic agent as a result
CC of M1, and the data is sufficient to convey the chemical structure and/or
CC properties of the agent. M1 can be used in the treatment of cancer such
CC as colon, breast, stomach, lung, thyroid, oesophageal, ovarian, kidney,
CC prostate or pancreatic cancer, adenocarcinoma, carcinoma, clear cell
CC cancer, infiltrating ductal cancer, infiltrating lobular cancer, squamous
CC cell carcinoma, neuroendocrine carcinoma, papillary carcinoma and Wilm's
CC tumour
XX Sequence 41936 BP; 9311 A; 11171 C; 11849 G; 9605 T; 0 U; 0 Other;
SQ
Query Match 34.8%; Score 313.6; DB 6; Length 41936;
Best Local Similarity 62.4%; Pred. No. 1.3e-58;
Matches 545; Conservative 0; Mismatches 319; Indels 10; Gaps 3;
QY 29 TGGGCACAGTGACTCAAGCCTGTACCAAGCAGCTTTCGAACTCCAAGTGGTGCATCA 88
Db 1336 TGGGCACGCTGGCTCAGCTCTGTAATCTCAGCACCTTGGAGGCAAGTGGTGCATCA 1395
QY 89 CTTGAGGTGAGAAAGTTTCGAGACCAAGCAGCTGGTCAATATGGTGAACCCCTATCTCTATAA 148
Db 1396 CTTGAGGTGAGAAAGTTTCGAGACCAAGCAGCTGGTCAATATGGTGAACCCCTCTCTATAA 1455
QY 149 AATACAAAATTAGTGGGTGTAGTGTAGTGCATGCTGTAGTCCAGCTACTCTGGGAGGCT 208
Db 1456 AATACAAAATTAGTGGGTGTAGTGTAGTGCATGCTGTAGTCCAGCTACTCTGGGAGGCT 1515

QY	209	GAGGCAAGAGAAATGCTTGAACCTGGGAGCGAGAGGTTGCAGTGAGCCGAGATGCCACCA	268	FT	/*tag= a	/standard_name= "Single nucleotide polymorphism"
Db	1516	GAGGCATGAGAAATCGCTTGAACCCAGAGCGGAGAGGTTGCAGTGAGCTGAGATCGCACCA	1575	FT	/complement(532)	
QY	269	CTGCACTCCAGCTGGGCGACACAGCGAGACTATCTCAAAAAATAAATAAATAAT	328	FT	/*tag= b	/standard_name= "Single nucleotide polymorphism"
Db	1576	CTGTACTCAGTCTCGGGCGACAGAGTGAGACTCTGCTCTCAAAAAAAAAAAAAAAAAA	1635	FT	/*tag= c	/complement(1100)
QY	329	AAAGGATCGGAGAGAAACAAAATAAATAAGATTCTCGGAAGGTGAACGAGATACGTAAT	388	FT	/*tag= d	/standard_name= "Single nucleotide polymorphism"
Db	1636	GACTCGTCAAGGTATAGATGTGACAGAGTACTAAGTGTTCGAAG-----AAAA	1687	FT	/*tag= e	/complement(1171)
QY	389	TATATGTAATAAGTTTAAATGCAATTTTAACTGTAAATCTTATTTTGGTTATAA	448	FT	/*tag= f	/standard_name= "Single nucleotide polymorphism"
Db	1688	TAAACACAGGCTGGGTGCATTTGGCTCATGCTGTAAATTTTCAGCACTTTGGGAGGCCAAG	1747	FT	/*tag= g	/complement(1351)
QY	449	AAGTAAACAGCCAAAGTAATGCACTTCAAACTCTACATAATATCTATNTATGGAAG	508	FT	/*tag= h	/standard_name= "Single nucleotide polymorphism"
Db	1748	GCAGGAGGATCACTTGAGCCCTAGGAGTTTGAGACCCAGCTCGACACAAATGAGACCCC	1807	FT	/*tag= i	/complement(1702)
QY	509	TGGAGGATCTATAATCCTACTACCCAAAGATAACCAAGTTTACATATTCCTCCAGATTTT	568	FT	/*tag= j	/standard_name= "Single nucleotide polymorphism"
Db	1808	ATGTTCTACAAAATTTTAAATAATTTAAATAATAGCTGGGCATGTGGCATGTGCTGTGG	1867	FT	/*tag= k	/complement(1860)
QY	569	TGGGCGATACACTAGCTTTTTTTTATTTGGGAAAAATTTCCATGTGCAGGCATACCTAATTT	628	FT	/*tag= l	/standard_name= "Single nucleotide polymorphism"
Db	1868	TCCCGCTGCTCAGGAGCTCAGGTGGGAGGATGCTTGGGCTTGAGAGGTCAAGGCTTC	1927	FT	/*tag= m	/complement(2162)
QY	629	TTCTAATGTCTNTAGTATTCATTTAAGGATGTTCCATAAATTTTAAATAATCATGCT	688	FT	/*tag= n	/standard_name= "Single nucleotide polymorphism"
Db	1928	AGTGAGTCATGATCGTGCCACTGCAATTCAGCCTGGGTGA-CAGAGTGAGACCCCTGCTTT	1986	FT	/*tag= o	/complement(2238)
QY	689	TTAAAGTAGAGAACTAGTGTGGGATGTCGCTCAGCCCTGT-ATCCGACACTTTGGG	747	FT	/*tag= p	/standard_name= "Single nucleotide polymorphism"
Db	1987	GAAATGAAGAAATAGCTTGGGCGAGTGGCGCATGGCTCACACCTGTATCCAGCACTTTGGG	2046	FT	/*tag= q	/complement(2373)
QY	748	AGGCGGAGGCAAAATGGATCACTTCAGGTCGCGAGTTTGAGACCAAGCTCGGCAACATGAT	807	FT	/*tag= r	/standard_name= "Single nucleotide polymorphism"
Db	2047	AGGCGGAGTGGTGGATCACTCAGGTTCAGAGATCGAGACCAAGCTGGCCAAATGGT	2106	FT	/*tag= s	/complement(2702)
QY	808	GAACACCCCTCTTAATAAATAACAAAAATAGCTGGGCAATGTGGCAAGCACTCTAG	867	FT	/*tag= t	/standard_name= "Single nucleotide polymorphism"
Db	2107	GAAATCCCATCTCTACTAAAAATACAAAAATTTAGCGGCGTGGTGGTGGCGGCTGTAA	2166	FT	/*tag= u	/complement(24778)
QY	868	TCCGAGCTACTCAGGAGTCTCAGGCGAGGATATC	901	FT	/*tag= v	/standard_name= "Single nucleotide polymorphism"
Db	2167	TCCGAGCTACTCGGAGGCTTGAGGAGGAGAAATC	2200	FT	/*tag= w	/complement(24830)
RESULT 9				FT	/*tag= x	/standard_name= "Single nucleotide polymorphism"
ADT25932				FT	/*tag= y	/complement(24840)
ID	ADT25932	standard; DNA; 56580 BP.		FT	/*tag= z	/standard_name= "Single nucleotide polymorphism"
XX	AC	ADT25932;		FT	/*tag= aa	/complement(3166)
XX	XX	30-DEC-2004 (first entry)		FT	/*tag= ab	/standard_name= "Single nucleotide polymorphism"
DT	DE	Wild type human ATP-gated ion channel P2X7R genomic DNA.		FT	/*tag= ac	/complement(24778)
XX	KW	antidepressant; tranquilizer; antimanic; neuroleptic;		FT	/*tag= ad	/standard_name= "Single nucleotide polymorphism"
KW	KW	ATP-gated ion channel P2X7R modulator; P2X7R; ATP-gated ion channel;		FT	/*tag= ae	/complement(24841)
KW	KW	affective disorder; depression; dysthymia; atypical depression;		FT	/*tag= af	/standard_name= "Single nucleotide polymorphism"
KW	KW	premenstrual dysphoric disorder; seasonal affective disorder;		FT	/*tag= ag	/complement(24942)
KW	KW	anxiety disorder; panic disorder; phobia; agoraphobia; social phobia;		FT	/*tag= ah	/standard_name= "Single nucleotide polymorphism"
KW	KW	obsessive-compulsive disorder; post-traumatic stress disorder;		FT	/*tag= ai	/complement(25010)
KW	KW	separation anxiety disorder; mania; hypomania; cyclothymic disorder;		FT	/*tag= aj	/standard_name= "Single nucleotide polymorphism"
KW	KW	bipolar disorder; human; gene; de; SNP; single nucleotide polymorphism.		FT	/*tag= ak	/complement(25010)
OS	XX	Homo sapiens.		FT	/*tag= al	/standard_name= "Single nucleotide polymorphism"
XX	XX	Key		FT	/*tag= am	/complement(26134)
XX	XX	Location/Qualifiers		FT	/*tag= an	/standard_name= "Single nucleotide polymorphism"
FT	FT	variation		FT	/*tag= ao	/complement(26202)
FT	FT	complement(362)		FT	/*tag= ap	/standard_name= "Single nucleotide polymorphism"

QY	522	TAATCCTACTACCCAAAGATAACAGTTACATATTTCTCCAGATTTTGGGGCATACACT	581	exon	/number= 3
Db	49116	TTGTACCAGTGGCTGCTTCTCGAAAGGAGCTTAGGGTGGGGAGTAGGAGTAGTCAAGA	49175	intron	/*tag= h
QY	582	AGCTTTTTTTATTTGGGAAAAATTTCCATGTGCAGGCATACCTAATTTTCTAAATGTCCTA	641	exon	/number= 4
Db	49176	GATTTTAGCGTCACTGTAATGTTTGTGATTGATTAATTCAGACTTTATCAAGCAGGTCC	49235	intron	/*tag= i
QY	642	TGTAGTATTCATTTAAGGATGTTCCATAATTTTTTAAATAACATGCTTTTAAAGTAGAGAA	701	exon	/number= 5
Db	49236	TCTGCGTTCAACTCCATGATTTTCCCAAGATAAATCTCTGGTACCTAAAAACAAAGAC	49295	intron	/*tag= j
QY	702	ACTAGGTGGGCATGTTGGCTCAGCCCTGTA-TCCGAGCACTTTGGGAGGCCGAGGCAAA	760	exon	/number= 6
Db	49296	GATTTGGCTAGACGTTGGTGGCTCAGCCCTGTAGTCCCACTTTTGGGAGGCCAAGGAGGG	49355	intron	/*tag= k
QY	761	TGGATCACTTGAAGTCCGGAGTTTGGACACGAGCTGGACACATGATGAACACCCCTCTC	820	exon	/number= 7
Db	49356	TGGATCACTGAGTCAAGGATTTCAAAACCAAGCTGGCCAACTGGGAGAAACCCCGTCTC	49415	intron	/*tag= l
QY	821	TAATAAAAATAC-AAAAATTAGCTGGGCGATGGTGGCAAGCACCTGTAGTCCCGACTACTC	879	exon	/number= 8
Db	49416	TACTAAAAATACAAAAAATTAGCCAGGCATGGTGGCGCATGCCCTGTATCCAGCTTCTC	49475	intron	/*tag= m
QY	880	AGGAGTCTGAGCGAGGAGTATC	901	exon	/number= 9
Db	49476	TAGAGGCTGAGCGAGGAGATC	49497	intron	/*tag= n
RESULT 10					
ADU23759					
ID	ADU23759 standard; DNA; 56580 BP.				
XX					
AC	ADU23759;				
XX					
DT	13-JAN-2005 (first entry)				
XX					
DE	Human ATP-gated ion channel protein (P2X7R) gene - SEQ ID 1.				
XX					
KW	ATP-gated ion channel protein; P2X7R; affective disorder;				
KW	major depression; dysthymia; atypical depression;				
KW	premenstrual dysphoric disorder; seasonal affective disorder;				
KW	generalised anxiety disorder; panic disorder; phobia; agoraphobia;				
KW	social phobia; specific phobia; obsessive-compulsive disorder;				
KW	post-traumatic stress disorder; separation anxiety disorder; mania;				
KW	hypomania; cyclothymic disorder; type I bipolar disorder;				
KW	type II bipolar disorder; gene; ds.				
OS	Homo sapiens.				
XX					
FH	Key				
FT	CDS				
FT	Location/Qualifiers				
FT	/*tag= b				
FT	/product= "Human ATP-gated ion channel protein (P2X7R) -				
FT	SEQ ID 10"				
FT	exon				
FT	3000. .3124				
FT	/*tag= a				
FT	/number= 1				
FT	intron				
FT	3125. .24840				
FT	/*tag= c				
FT	/number= 1				
FT	exon				
FT	24841. .25009				
FT	/*tag= d				
FT	/number= 2				
FT	intron				
FT	25010. .26133				
FT	/*tag= e				
FT	/number= 2				
FT	exon				
FT	26134. .26202				
FT	/*tag= f				
FT	/number= 3				
FT	intron				
FT	26203. .30957				
FT	/*tag= g				
FT					

FT	exon	/number= 3
FT		30958. .31030
FT		/*tag= h
FT		/number= 4
FT	intron	31031. .32480
FT		/*tag= i
FT		/number= 4
FT	exon	32481. .32577
FT		/*tag= j
FT		/number= 5
FT	intron	32578. .35415
FT		/*tag= k
FT		/number= 5
FT	exon	35416. .35496
FT		/*tag= l
FT		/number= 6
FT	intron	35497. .36112
FT		/*tag= m
FT		/number= 6
FT	exon	36113. .36242
FT		/*tag= n
FT		/number= 7
FT	intron	36243. .37540
FT		/*tag= o
FT		/number= 7
FT	exon	37541. .37677
FT		/*tag= p
FT		/number= 8
FT	intron	37678. .45469
FT		/*tag= q
FT		/number= 8
FT	exon	45470. .45560
FT		/*tag= r
FT		/number= 9
FT	intron	45561. .47229
FT		/*tag= s
FT		/number= 9
FT	exon	47230. .47295
FT		/*tag= t
FT		/number= 10
FT	intron	47296. .47379
FT		/*tag= u
FT		/number= 10
FT	exon	47380. .47529
FT		/*tag= v
FT		/number= 11
FT	intron	47530. .50437
FT		/*tag= w
FT		/number= 11
FT	exon	50438. .50539
FT		/*tag= x
FT		/number= 12
FT	intron	50540. .54391
FT		/*tag= y
FT		/number= 12
FT	exon	54392. .54889
FT		/*tag= z
FT		/number= 13
XX		
PN	EP1473367-A1.	
XX		
PD	03-NOV-2004.	
XX		
PF	16-APR-2004; 2004EP-00018640.	
XX		
PR	17-APR-2003; 2003EP-00008753.	
PR	04-SEP-2003; 2003EP-00019626.	
XX	16-APR-2004; 2004EP-00009160.	
XX		
PA	(NEUR-) NEURONOVA AG.	
XX		
PI	Barden N, Sillaber I, Paez-Pereda M;	
FT		

DR WPI: 2004-787080/78.
DR P-PSDB; ADU23768.
XX
PT Use of nucleic acid molecules comprising sequence encoding P2X7R with a
PT mutation in 5'UTR, 3'UTR, exon 3, 5, 6, 8 or 13, introns 1, 3, 4 - 7, 9,
PT 11 or 12 or deletion of exon 13 for preparation of composition for
PT detecting affective disorder.
XX
PS Claim 1; SEQ ID NO 1; 188pp; English.
PS
XX The invention comprises the amino acid and coding sequences of a human
CC ATP-gated ion channel protein (P2X7R), the invention also comprises
CC allele-specific PCR primers for the human P2X7R gene. The DNA and protein
CC sequences of the invention are useful in the preparation of an affective
CC disorder, such as: major depression (e.g. dysthymia, atypical depression,
CC premenstrual dysphoric disorder and seasonal affective disorder),
CC generalised anxiety disorder (e.g. panic disorder, phobias, agoraphobia,
CC social phobia, specific phobia, obsessive-compulsive disorder, post-
CC traumatic stress disorder, separation anxiety disorder, mania, hypomania
CC and cyclothymic disorder), and bipolar disorder (e.g. type I or type II
CC bipolar disorder). The present DNA sequence represents the human P2X7R
CC gene of the invention.
XX
SQ Sequence 56580 BP; 15379 A; 13163 C; 13083 G; 14955 T; 0 U; 0 Other;
Query Match 34.8%; Score 313.6; DB 13; Length 56580;
Best Local Similarity 62.3%; Pred. No. 1.4e-58;
Matches 574; Conservative 0; Mismatches 319; Indels 29; Gaps 4;
QY 9 TTAATTAATAAATCACTCAGTTCGGGCACAGTGAAGCTGTAACACAGACTTTGGA 68
DB 48576 TTAATAATTTGGCCCAAGCTGGGCACAGTGTGCTCAGGCTATATATCCCAACACTTTGG 48635
QY 69 AGTCCAAAGTGGGTGATCACTTGTAGTGAAGAAGTTCGAGACAGCCTGGTCAATATGTT 128
DB 48636 AGCCGAGGCGAGCAGTCACTTGTAGTCAAGGTTCGAGACAGCCTGGCCCAACATGTT 48695
QY 129 GAAACCTTCTCTACTAAAAATACAAAAATTAAGCTGGGTGTAGTGTGATGCTGCTGTAG 188
DB 48696 GAAACCCCTCTCTACTAAAAATACAAAAATTAAGCCAGCGTGGTGTGATGCTGCTGAAG 48755
QY 189 TCCAGCTACTCGGAGGCTGAGGCAAGAGAAATTCCTTCAACTGGGAGGAGAGGTTGC 248
DB 48756 TCCAGCTACTTTGGGAGGCTGAGGCAAGAGAAATTCCTTCAACTGGTGTAGGAGGTTGC 48815
QY 249 AGTGAGCGAGATCCACCACTGCTCCTCCAGCTGGGCGACACAGCAGACTCTATCTCA 308
DB 48816 AGTGAGCTGAGATCGCCATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTCA 48875
QY 309 AAAAAATAAATAAATAAAGGATCGGAGAGAAACAAACTAATAAGATTCCTCTGAAG 368
DB 48876 AAAAAATAAATAAATAAAGGATCGGAGAGAAACAAACTAATAAGATTCCTCTGAAG 48935
QY 369 GTAA-----GCAGAGATACGTAAATTTATATATATATATATATATATATATATATAT 411
DB 48936 TTAATAAGAAAAAGCAAACTGTAGAACATATATATATATATATATATATATATATATAT 48995
QY 412 ATTTTA-----ACTGTAATCTATTGTTTATTTTGGTTATATAAAGTAAACAGCC 461
DB 48996 AAATCACTACACAGAGAGTATCTGTATATTTTCTGATGTATATCTATATATACAC 49055
QY 462 AAAAGTAATGCAACTTTCAAACCTACATAAATATCTATTATGGAAGAGTGGAGGCATCTA 521
DB 49056 AAGGTACACACACACACATACATATATATATATATATATATATATATATATATATATAT 49115
QY 522 TAATCTACTACCAAGATAACCAAGTACATATCTCCAGATTTTTCGGGATACACT 581
DB 49116 TTGTACAGTGGCTGCTTCTGSAAGAGAGCTTAGGGTGGGGAGTAGGAGTAGTCAAGA 49175
QY 582 AGCTTTTATTTTGGGAAAAATTTCCATGTGCGAGGATACCTAATTTTCTTAATGTCTA 641
DB 49176 GATTTAGCGTCACTGCTATTTGTTGATTTGATTAATTCAGACTTTATCAACAGGTC 49235

QY 642 TGTAATTCATTTAAGGATGTTCCATAATTTTAAAAATACATATGCTTTAAAGTAGAGAA 701
DB 49236 TCTCGGTTCACTCCATGATTTTCCCAAGATATAATCTCTGGTACCTAAAAACAAGAC 49295
QY 702 ACTAGGTTGGGCATGGTGGCTCAGCCTGTA-TCCACAGCACTTTGGAGGCCGAGGCAAA 760
DB 49296 GATTGGCTAGACGCTGGTGGCTCAGCCTGTAAGTCTCCAGTACTTTGGGAGGCCAAGGAGG 49355
QY 761 TGGATCACTTGGATCCGGAGTTTGACACAGCCTGCAACATGATGAACACCCCTCTC 820
DB 49356 TGGATCACTTGGATCCGGAGTTTCAAAAACCCAGCCTGGCCCAACATGGGAGAAACCCGCTC 49415
QY 821 TAATAAAAAATAC-AAAAATTTAGCTGGGCATGTTGGCAAGCACCTGTAGTCCCACTACTC 879
DB 49416 TACTAAAAAATACAAAAAATTTAGCCAGGCATGTTGGCCATGCTGCTGTATATCCACGTTCTC 49475
QY 880 AGGAGTCTGAGGCGAGGAGTATC 901
DB 49476 TAGAGGCTGAGGCGAGGAGATC 49497
RESULT 11
AA83427
ID AAX83427 standard; DNA; 21721 BP.
XX
AC AAX83427;
XX
DT 31-AUG-1999 (first entry)
XX
DE Human lipolysis stimulated receptor genomic sequence.
XX
KW Human; lipolysis stimulated receptor; LSR; lipoprotein; free fatty acid;
KW cytokine; probe; primer; amplification; hybridisation; detection; leptin;
KW allele; variant; mutation; deletion; loss of heterozygosity; chylomicron;
KW transgenic animal; gene expression; triglyceride; eating disorder;
KW obesity; atheromatosis; atherosclerosis; hypertension; diabetes;
KW anorexia; metabolism; ds.
XX
OS Homo sapiens.
XX
PN WO9907737-A2.
XX
PD 18-FEB-1999.
XX
PF 06-AUG-1998; 98WO-IB001257.
XX
PR 06-AUG-1997; 97FR-00010088.
XX
PR 22-APR-1998; 98FR-00005032.
XX
XX (GEST) GENSET.
PA (INRM) INSERM INST NAT SANTE & RECH MEDICALE.
XX
XX Bihaïn B, Bougueleret L, Yen-Potin F;
XX
XX WPI; 1999-190035/16.
DR
XX
XX Lipolysis stimulated receptor involved in leptin metabolism - and
PT controlling its activity for treatment of obesity, anorexia or abnormal
PT cytokine metabolism.
XX
XX Claim 40; Page 258-277; 279pp; French.
PS
XX This sequence represents the human lipolysis stimulated receptor (LSR)
CC genomic sequence. The novel LSR binds lipoproteins in presence of free
CC fatty acids (FFA) and binds cytokines in absence of FFA. The nucleic acid
CC and its fragments are used as probes and primers for detection and/or
CC amplification of LSR genes; for production of recombinant LSR; for
CC detecting allelic variants, mutations, deletions, loss of heterozygosity
CC and genetic abnormalities in the gene. LSR, recombinant cells and
CC transgenic animals are used to screen for chemical interactions with LSR,
CC also to study expression and activity of LSR and its interactions. The
CC chemicals, and leptin, are used to modulate the number of LSR in a cell,
CC its recycling rate and/or specificity of receptor activity, particularly

CC for reducing the level of leptin, lipoproteins, chylomicrons and/or triglycerides. The chemicals are thus useful for treating eating disorders, particularly obesity (and related diseases such as atheromatosis, atherosclerosis, hypertension and diabetes) or anorexia, also disease associated with abnormal cytokine metabolism	XX	Sequence 21721 BP; 4980 A; 5688 C; 6170 G; 4865 T; 0 U; 18 Other;
Query Match	34.7%; Score 312.6; DB 2; Length 21721;	
Best Local Similarity	62.2%; Pred. No. 1.9e-58;	
Matches 544; Conservative	0; Mismatches 320; Indels 10; Gaps 3;	
Qy 29	TGGGCACAGTCACTCAAGCCTGTAAACCACTACTTTTGGAAAGTCCAAAGGTGGGTGGATCA	88
Db 5494	TGGGCACGGTGGCTCAGCTCTGTAACTCTCAGCACTTTGGGAGGCCAAGGTGGGTGGATCA	5553
Qy 89	CTTCAGGTGACAAGTTTCGAGACACAGCCTGGTCAATATGTTGTAACCCCTATCTCTACTAAA	148
Db 5554	CCTCAGGTTCAGAGTTTCAGACCAAGCCTGGCCACATGTTGGTGAACCCCGTCTCTACTAAA	5613
Qy 149	AATACAAAAATTTAGCTGGGTGTAGTGTGATGCTGTGTAGTCCAGCTACTCTGGGAGGCT	208
Db 5614	AATACAAAGATTAGCCAGGCATGTTGGCCATGCTGTATCTCCAGCTACTCAGAGGCT	5673
Qy 209	GAGCAAGAGAAATTTGCTTGAACCTGGGAGGCAGAGTTTGCAGTGAGCGCAGATCCCAACCA	268
Db 5674	GAGCATGAGAAATCGCTTTGAACCCAGGAGGCAGAGTTTGCAGTGAGCTGAGATCGCACCA	5733
Qy 269	CTGCACCTCCAGCCTGGGGGACACAGCGAGCTCTATCTCAAAAAATAAATAAATAAAT	328
Db 5734	CTGTACTTGCAGTTNTGGCGCAGAGTGCAGACTCTGTCTCAAAAAATAAATAAATAAAT	5793
Qy 329	AAAGGATCGGAGAGAAACAAACTAATAAGATTCCTGAAGGTAGCAGATACGTAAAT	388
Db 5794	GATCCCTCAGGTATAGAAATGTCAGAGTACTAAAGTTGTCAAAG-----AAAA	5845
Qy 389	TATATGTAATAAAGTTTTAAATGCAATTTTAACTGTAACTCTTATTTTATTTTGGTTATAA	448
Db 5846	TACACACAGGCTGGGTGCATTTGGCTCATGCTGTAAATTTTTCAGCCTTTGGGAGGCCAAG	5905
Qy 449	AAGTAAACAGGCCAAAAAGTAAATGCAACTTCAAACTCTACATAATATCTATTATGGAAG	508
Db 5906	GCAGGAGGATCACTTGAGCGCTTAGAGTTTGAGACCAAGCTGGCAACAAATAGAGACCCC	5965
Qy 509	TGGNAGCATCTAATAATCCTACTACCAAGATAACAGTTACATATTCCTCCAGATTTT	568
Db 5966	ATGTCTCAAAAAATTTTAAAAAATTTAAATAATAGCTGGGCATGTGCCTGTGG	6025
Qy 569	TGGGGCATACACTAGCTTTTTTTTATTTGGGAAAAATTTCCATGTGCAGGCATACCTAATTT	628
Db 6026	TCCCGGTGCTCAGGAGCTGAGGTGGAGGATTTGCTTGGGCTTGAGAGTCAAGGCTTC	6085
Qy 629	TTCTAAATGTCTATGTAGTATTCCATTTAAGGATGTTTCCAATAATTTTAAAAATACATGCT	688
Db 6086	AGTGAGTCAATGATGCTGCCACTGCATTTCCAGCCTGGGTGA-CAGAGTGAGACCTGTCTT	6144
Qy 689	TTAAAGTAGAGAAACTAGTTTGGGCATGTTGGCTCAGCCTGT-ATCCCAGCACTTTGGG	747
Db 6145	GAATGAAAAAGAAAAATAGGCTGGCGCAGTTGGCTCACACCTGTAATCCCAGCACTTTGGG	6204
Qy 748	AGGCCGAGGCAATGGATCACTTTGAGTCCGGAGTTTGAGACCAAGCCTGGACAACATGAT	807
Db 6205	AGGCCGAGGTGGGTGGATCACTTACCTGAGGTCAAGGATCGAGACCAAGCCTGGCAACATG	6264
Qy 808	GAACAACCCCTCTCTAATAAAAAATACAAAAATTAAGCTGGGCATGTTGGTGAAGCAACCTGTAG	867
Db 6265	GAATCCCATCTCTACTAATAATAAATAAATTTAGCCGGGGCTGGTGGTGGCGCTGTAA	6324
Qy 868	TCCAGGCTACTCAGGAGTCTGAGCGAGGAGTATC	901
Db 6325	TCCAGGCTACTCAGGAGGCTGAGCGAGGAGTATC	6358

RESULT 12	
ADSI9371/c	
ID	ADSI9371 standard; DNA; 37940 BP.
XX	
AC	ADSI9371;
XX	
DT	18-NOV-2004 (first entry)
XX	
DE	Human PIP5K type II beta related DNA #1.
XX	
KW	phosphatidylinositol-4-phosphate 5-kinase type II-beta; PIP5KII-beta;
KW	hyperproliferative disorder; cancer; inflammation; metabolic disorder;
KW	angiogenesis; therapy; human; ds.
XX	
OS	Homo sapiens.
XX	
PN	US2003232777-A1.
XX	
PD	18-DEC-2003.
XX	
PP	16-JAN-2003; 2003US-00348073.
XX	
PR	18-JUN-2002; 2002US-00175627.
XX	
PA	(MARC/) MARCUSSON E G.
PA	(DOBI/) DOBIE K W.
PA	(FREI/) FREIER S M.
XX	
PI	Marquesson EG, Dobie KW, Freier SM;
XX	
DR	WPI; 2004-070378/07.
DR	GENBANK; AC006449.19.
XX	
PT	New compound targeted to a nucleic acid molecule encoding
PT	phosphatidylinositol-4-phosphate 5-kinase, type II beta, useful in
PT	treating hyperproliferative disorder, e.g. cancer, inflammatory or
PT	metabolic disorder.
XX	
PS	Example 13; SEQ ID NO 12; 67pp; English.
XX	
CC	The present invention is directed to antisense oligonucleotides which are
CC	targeted to a nucleic acid encoding phosphatidylinositol-4-phosphate 5-
CC	kinase type II (PIP5KII) beta and which modulate the expression of
CC	PIP5KII. The invention is useful for treating a disease or condition
CC	associated with phosphatidylinositol-4-phosphate 5-kinase type II beta
CC	such as hyperproliferative disorder (cancer), inflammatory, metabolic
CC	disorder and inhibiting angiogenesis. The present sequence is human
CC	PIP5KII related DNA.
XX	
SQ	Sequence 37940 BP; 8915 A; 9323 C; 9359 G; 10343 T; 0 U; 0 Other;
Query Match	34.7%; Score 312.4; DB 13; Length 37940;
Best Local Similarity	64.3%; Pred. No. 2.4e-59;
Matches 561; Conservative	0; Mismatches 281; Indels 30; Gaps 5;
Qy	31 GGCACAGTACTCAAGCCTCTAAACCACAGTACTTTGGAAGTCCAAAGTGGGTGGATCACT 90
Db	5175 GGTGAGTGGCTCAGCGCTGTATCCACACCTTTGGAGGCGGAGCGGGGGATTGCT 5116
Qy	91 TGAGGTGAGAAGTTCCAGACCCAGCCTGGTCAATATGGTGAACCCTATCTCTACTAAAAA 150
Db	5115 TGAGGCCAGGAGTTCGAGACCCAGTCTGGCCAAACATGSCAAACCCCTGCTCTACTAAAAA 5056
Qy	151 TACAAAANTTAGCTGGGTGTAGTGATGCATGCTGTAGTCCAGCTACTCGGAGGCTGA 210
Db	5055 TACAAAANTTAGCTGGTGTGGTGGCGGGCACCTGTAATCCAGTTACTCAGGAGGCTGA 4996
Qy	211 GGCAAGAGAATTCCTTGAACCTGGGAGGCGAGAGTTCAGTGTAGCCGAGATCCCAACCACT 270
Db	4995 GGCAAGAGAATTCCTTGAACCCGGGAGGCGAGAGTTCAGTGTAGCCGAGATCCCAACCACT 4936
Qy	271 GCATCTCAGCCTGGGCGACACAGCGAGCTCTATCTCAAAAAAATAAATAAATAAATAA 330

Db 4935 GCACCTCCAGCCGCGGAGCAGAGCAAACTCCATCTCAAAAAAAAAAAAAAAAAAAAAA 4876

Qy 331 AGGATCGGAGAGAAACAACTAAATAGATTCCTGAAGTAAAGCAGAGATACGTAAATTA 390

Db 4875 AAGAAAATGAAAGAGACAACCTCA-----ATGAATAGGATAAAATATTGCAAAATCA 4824

Qy 391 TATGTAATAAGTTTAAATGCAATTTAACTGTAATCTTATGTTTATTTTGGTTATAAAA 450

Db 4823 TATATCTGATAGGGACCTGTATCTAGAAATATAAAAGAAATTCCTTGT-----AA 4775

Qy 451 GTAAACAGCCAAAAGTAAATGCACTCAAACTCTACATAAATATCTATTATGGAAGTG 510

Db 4774 CTCATAATAGAGATACCAATTTAATATGGGCAAAAGATC---TTTGTAATGG 4718

Qy 511 GAAGGCATCTATAATCTTACTACCAAGATAACCAAGTTACATATTCCTCCAGATTTTG 570

Db 4717 GCAAAAATTTAATAGACATTTCCAAAGAAAAGATACCAATGAAGGCTGGGATGTGCG 4658

Qy 571 GGGCATACACTAGCTTTTATTTTGGGNAATTTCCATGTCGAGGCATACCTAATTTT 630

Db 4657 TCATGCTGCAATCCCACTCTCTTGGGAGGCCAAGCGGGAGGA-----CCACTTGAG 4605

Qy 631 CTAAATGCTATGATGATTCATTTAAGGATGTTCCATAATTTTAAATACATGCTTT 690

Db 4604 GCAGAGTTCAGGACCGCTCGGCAACAAAGCGAGACCCCTGTCTCTACAAACAACA 4545

Qy 691 AAAGTAGAGAACTAGTTGGGCATGGTGGCTCAGCCCTGT-ATCCAGCACTTTGGGAG 749

Db 4544 AAAAGAAATGTACAGCCAGGCACAGTCGCTCACACCTGTATCCAGCACTCTGGGA 4485

Qy 750 GCGAGGCAAAATGATTCATCTAGAGTCCGAGTTTGAGACACAGCTGGACAACATGATGA 809

Db 4484 GCCAAGGCAGTGGATCAGCTCAGGTTAGGAGTTTGAGATCAGCCTGGACAACATGGTAA 4425

Qy 810 AACACCTCTCTAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATA 869

Db 4424 AACTCCATATCTACTAAGAAATACAAAGTTAGCTGGGTGGTGGCGACCTGTATC 4365

Qy 870 CCAGCTACTCAGGAGTCTCAGGCGAGGATATC 901

Db 4364 CGAGGTACTTGGAGGCTGAGCGAGGAGATC 4333

RESULT 13

ADJ26093/c

ID ADJ26093 standard; DNA; 38000 BP.

XX AC ADJ26093;

XX DT 20-MAY-2004 (first entry)

XX DE Human PIP5KII-beta genomic DNA SeqID 11.

XX KW human; PIP5KII-beta;

XX KW phosphatidylinositol 4-phosphate 5-kinase, type II beta;

XX KW cell differentiation; proliferation; motility; cancer; inflammatory;

XX KW metabolic disease; infection; tumour formation; cytostatic;

XX KW antiinflammatory; ds.

XX OS Homo sapiens.

XX PN US2003232775-A1.

XX PD 18-DEC-2003.

XX PF 18-JUN-2002; 2002US-00175627.

XX PR 18-JUN-2002; 2002US-00175627.

XX PA (ISIS-) ISIS PHARM INC.

XX PI Dobie KW, Freier SM;

XX XX

DR WPI; 2004-061311/06.

XX New antitense compound targeted to a nucleic acid molecule encoding

PT phosphatidylinositol-4-phosphate 5-kinase, type II, useful for modulating

PT expression of phosphatidylinositol-4-phosphate 5-kinase, type II or

PT treating cancer.

XX Example 15; SEQ ID NO 11; 63pp; English.

XX This invention relates to novel compositions and methods for modulating

CC the expression of PIP5KII-beta (also known as phosphatidylinositol 4-

CC phosphate 5-kinase, type II beta). Specifically, it refers to antitense

CC compounds that can target and hybridise with a nucleic acid molecule

CC encoding PIP5KII-beta, as well as splice variants thereof, to inhibit

CC expression accordingly. PIP5KII-beta is a homodimeric enzyme that is a

CC member of the superfamily of phosphatidylinositolphosphate kinases and is

CC thought to play a role in pathological conditions in which cell

CC differentiation, proliferation and/or motility are compromised, such as

CC cancer, inflammatory disorders and metabolic diseases. The present

CC invention describes antitense compounds that can be used as research

CC tools in differential and/or combinatorial analyses to elucidate

CC expression patterns of a portion or the entire complement of genes

CC expressed within cells and tissues. The compound can also be used as

CC prophylaxis, e.g. to prevent or delay infection, inflammation or tumour

CC formation. Accordingly, these compounds can be described as having

CC cytostatic and antiinflammatory activities. This polynucleotide sequence

CC is the human PIP5KII-beta genomic DNA of the invention.

XX

SQ Sequence 38000 BP; 8927 A; 9338 C; 9374 G; 10356 T; 0 U; 5 Other;

Query Match 34.7%; Score 312.4; DB 12; Length 38000;

Best Local Similarity 64.3%; Pred. No. 2.4e-58;

Matches 561; Conservative 0; Mismatches 281; Indels 30; Gaps 5;

Qy 31 GGCACAGTGACTCAAGCCTGTAACCAACAGTAGTACTTTGGAAAGTCCAAAGTGGGTGGATCACT 90

Db 5175 GGTGAGTGGCTCAGCCCTGTAATCCCACTTTGGGAGCCGAGGGGGGATGCT 5116

Qy 91 TGAGTGAGAGTTTCGACACCAAGCTGTCATATGTTGAAACCTTCTACTATAAAA 150

Db 5115 TGAGGCCAGGAGTTCGACACCAAGCTGTCGCAACATGGCAAAACCTTCTACTATAAAA 5056

Qy 151 TACAAAAATTAGCTGGGTGTAGTCATGCTCTAGTCCAGCTACTCGGAGGCTGA 210

Db 5055 TACAAAAATTAGCTGGGTGTAGTGGCGGCACCTGTAATCCAGTTACTCAGAGGCTGA 4996

Qy 211 GGCAGAGAGAAATGCTTGAACCTGGAGGCGAGAGTTGAGTGAGCGGAGATCCCACT 270

Db 4995 GGCAGGAGAAATCGCTTGAACCCGCGGAGGCGAGAGTTGCGAGTGAAGATCACACCGTT 4936

Qy 271 GCATCCAGCTGGGCGACACAGCGAGCTCTATCTCAAAAAATAAATAAATAAATAA 330

Db 4935 GCATCCAGCTGGGCGACAGAGCAAACTCCATCTCAAAAAATAAATAAATAAATAA 4876

Qy 331 AGGATCGGAGAGAAACAAACTAAATAGATTCCTGAAGTAAAGCAGAGATACGTAAATTA 390

Db 4875 AAGAAAATGAAAGAGACAACCTCA-----ATGAATAGGATAAAATATTGCAATCA 4824

Qy 391 TATGTAATAAGTTTAAATGCAATTTTAACTGTAATCTTATGTTTATTTTGGTTATAAAA 450

Db 4823 TATATCTGATAGGGACCTGTATCTAGAAATATAAAAGAAATTCCTGT-----AA 4775

Qy 451 GTAAACAGCCAAAAGTAAATGCAATTTCAAACTCTACATAAATATCTATTATGGAAGTG 510

Db 4774 CTCATAATAGAGATACCAATTTAATATGGGCAAAAGATC---TTTGTAATGG 4718

Qy 511 GAAGGCATCTATAATCTTACTACCAAGATAACCAAGTTACATATTCCTCCAGATTTTG 570

Db 4717 GCAAAAATTTAATAGACATTTCCAAAGAAAAGATACCAATGAAGGCTGGGATGTGCG 4658

Qy 571 GGGCATACACTAGCTTTTATTTTGGGNAATTTCCATGTCGAGGCATACCTAATTTT 630

Db 4657 TCATGCTGCAATCCCACTCTCTTGGGAGGCCAAGCGGGAGGA-----CCACTTGAG 4605

QY 631 CTAATGTCTATGATGATTTCCATTTAAGGATGTTCCATAAATTTTAAATATACATGCTTT 690
Db 4604 GCCAGGAGTTTCAGGACCGCTCGGCAACAAAGCGAGACCCCTGTCTCTACACACAAACA 4545
QY 691 AAGTAGAGAACTAGTGGGATGCTGCTCAGCCCTGT-ATCCAGCAGCTTTGGGAG 749
Db 4544 AAAAGAAATGATACAGCCGAGGACAGTGGCTCACACCTGTAAATCCAGCAGCTCTGGGAA 4485
QY 750 GCCAGGCAATAGGATCACTTGAGGTCCGGAGTTTGAGACCAACGCTGGACACATGATCA 809
Db 4484 GCCAGGAGTGGATCACTGAGTTAGAGTTGAGATGACGCTGGACACATGTTAA 4425
QY 810 AACACCTCTCTAATAAAAAATACAAAAATAGCTGGGCACTGGTGGCAAGCAGCTGTAGTC 869
Db 4424 AACTCCATATCTACTAAGAAATACAAAAGTTAGCTGGTGTGGTGGCAGCAGCTGTAAATC 4365
QY 870 CCAGCTACTCAGAGTCTGAGCAGGAGTATC 901
Db 4364 CCAGGTACTTGGAGGCTGAGCGAGGAATC 4333
RESULT 14
AAL35901/c
ID AAL35901 standard; DNA; 28818 BP.
XX
AC AAL35901;
DT
DT 08-JAN-2002 (first entry)
XX
DE Human musculoskeletal system related polynucleotide SEQ ID NO 2266.
KW Cytostatic; immunosuppressive; nootropic; neuroprotective; antiviral;
KW anti-allergic; hepatotropic; antidiabetic; anti-inflammatory; anti-ulcer;
KW vulnary; aniconvulsant; antibacterial; antifungal; antiparasitic;
KW cardiac; gene therapy; cancer; immune disorder; cardiovascular disorder;
KW neurological disease; infection; human; secreted protein;
KW musculoskeletal system; ds.
XX
OS Homo sapiens.
XX
PN WO200155367-A1.
XX
PD 02-AUG-2001.
XX
PF 17-JAN-2001; 2001WO-US001338.
XX
PR 31-JAN-2000; 2000US-0179065P.
PR 04-FEB-2000; 2000US-0180628P.
PR 24-FEB-2000; 2000US-0184664P.
PR 02-MAR-2000; 2000US-0186350P.
PR 16-MAR-2000; 2000US-0189874P.
PR 17-MAR-2000; 2000US-0190078P.
PR 18-APR-2000; 2000US-0198123P.
PR 19-MAY-2000; 2000US-0205515P.
PR 07-JUN-2000; 2000US-0209467P.
PR 28-JUN-2000; 2000US-0214886P.
PR 30-JUN-2000; 2000US-0215135P.
PR 07-JUL-2000; 2000US-0216647P.
PR 07-JUL-2000; 2000US-0216880P.
PR 11-JUL-2000; 2000US-0217487P.
PR 11-JUL-2000; 2000US-0217496P.
PR 14-JUL-2000; 2000US-0218290P.
PR 26-JUL-2000; 2000US-0220963P.
PR 26-JUL-2000; 2000US-0220964P.
PR 14-AUG-2000; 2000US-0224518P.
PR 14-AUG-2000; 2000US-0225213P.
PR 14-AUG-2000; 2000US-0225213P.
PR 14-AUG-2000; 2000US-0225214P.
PR 14-AUG-2000; 2000US-0225266P.
PR 14-AUG-2000; 2000US-0225267P.
PR 14-AUG-2000; 2000US-0225268P.
PR 14-AUG-2000; 2000US-0225270P.
PR 14-AUG-2000; 2000US-0225447P.
PR 14-AUG-2000; 2000US-0225757P.
PR 14-AUG-2000; 2000US-0225758P.
PR 14-AUG-2000; 2000US-0225759P.
PR 18-AUG-2000; 2000US-0226279P.
PR 22-AUG-2000; 2000US-0226681P.
PR 22-AUG-2000; 2000US-0226688P.
PR 22-AUG-2000; 2000US-0227182P.
PR 23-AUG-2000; 2000US-0227009P.
PR 30-AUG-2000; 2000US-0228924P.
PR 01-SEP-2000; 2000US-0229287P.
PR 01-SEP-2000; 2000US-0229343P.
PR 01-SEP-2000; 2000US-0229344P.
PR 05-SEP-2000; 2000US-0229345P.
PR 05-SEP-2000; 2000US-0229509P.
PR 05-SEP-2000; 2000US-0229513P.
PR 06-SEP-2000; 2000US-0230437P.
PR 06-SEP-2000; 2000US-0230438P.
PR 08-SEP-2000; 2000US-0231242P.
PR 08-SEP-2000; 2000US-0231243P.
PR 08-SEP-2000; 2000US-0231244P.
PR 08-SEP-2000; 2000US-0231413P.
PR 08-SEP-2000; 2000US-0231414P.
PR 08-SEP-2000; 2000US-0232080P.
PR 08-SEP-2000; 2000US-0232081P.
PR 12-SEP-2000; 2000US-0231968P.
PR 14-SEP-2000; 2000US-0232397P.
PR 14-SEP-2000; 2000US-0232398P.
PR 14-SEP-2000; 2000US-0232399P.
PR 14-SEP-2000; 2000US-0232400P.
PR 14-SEP-2000; 2000US-0232401P.
PR 14-SEP-2000; 2000US-0233063P.
PR 14-SEP-2000; 2000US-0233064P.
PR 14-SEP-2000; 2000US-0233065P.
PR 21-SEP-2000; 2000US-0234223P.
PR 21-SEP-2000; 2000US-0234274P.
PR 25-SEP-2000; 2000US-0234977P.
PR 25-SEP-2000; 2000US-0234988P.
PR 26-SEP-2000; 2000US-0235484P.
PR 27-SEP-2000; 2000US-0235834P.
PR 27-SEP-2000; 2000US-0235836P.
PR 29-SEP-2000; 2000US-0236327P.
PR 29-SEP-2000; 2000US-0236367P.
PR 29-SEP-2000; 2000US-0236368P.
PR 29-SEP-2000; 2000US-0236369P.
PR 29-SEP-2000; 2000US-0236370P.
PR 02-OCT-2000; 2000US-0236802P.
PR 02-OCT-2000; 2000US-0237037P.
PR 02-OCT-2000; 2000US-0237038P.
PR 02-OCT-2000; 2000US-0237039P.
PR 02-OCT-2000; 2000US-0237040P.
PR 13-OCT-2000; 2000US-0239935P.
PR 13-OCT-2000; 2000US-0239937P.
PR 20-OCT-2000; 2000US-0240960P.
PR 20-OCT-2000; 2000US-0241221P.
PR 20-OCT-2000; 2000US-0241785P.
PR 20-OCT-2000; 2000US-0241786P.
PR 20-OCT-2000; 2000US-0241787P.
PR 20-OCT-2000; 2000US-0241808P.
PR 20-OCT-2000; 2000US-0241809P.
PR 01-NOV-2000; 2000US-0244617P.
PR 01-NOV-2000; 2000US-0244647P.
PR 08-NOV-2000; 2000US-0246475P.
PR 08-NOV-2000; 2000US-0246476P.
PR 08-NOV-2000; 2000US-0246477P.
PR 08-NOV-2000; 2000US-0246478P.
PR 08-NOV-2000; 2000US-0246523P.
PR 08-NOV-2000; 2000US-0246524P.
PR 08-NOV-2000; 2000US-0246525P.
PR 08-NOV-2000; 2000US-0246526P.
PR 08-NOV-2000; 2000US-0246527P.
PR 08-NOV-2000; 2000US-0246528P.

GenCore version 5.1.7
Copyright (c) 1993 - 2006 Bioacceleration Ltd.

OM nucleic - nucleic search, using sw model

Run on: February 9, 2006, 15:02:28 ; Search time 3367.06 Seconds
(without alignments)
12519.885 Million cell updates/sec

Title: US-10-607-806-1-T9182_COPY_8700_9600

Perfect score: 901

Sequence: 1 tggtcattataataaaat.....gagtcctgagcaggagatc 901

Scoring table: IDENTITY NUC

Gapop 10_0, Gapext 1.0

Searched: 41078325 seqs, 23393541228 residues

Total number of hits satisfying chosen parameters: 82156650

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : EST:

1: gb_est1:*
2: gb_est2:*
3: gb_est3:*
4: gb_hc:*
5: gb_est4:*
6: gb_est5:*
7: gb_est6:*
8: gb_est7:*
9: gb_ges1:*
10: gb_ges2:*
11: gb_ges3:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB	ID	Description
C 1	310.4	34.5	507	5	BU785040	BU785040 in43g09.x
C 2	281	31.2	2213	4	CR623485	CR623485 full-leng
C 3	280	31.1	1218	4	CR626853	CR626853 full-leng
C 4	276	30.6	6056	4	CR627178	CR627178 Homo sapi
C 5	260.4	28.9	903	5	BU431110	BU431110 BX431110
C 6	253.2	28.1	2495	4	HSM808204	BM648058 Homo sapi
C 7	253	28.1	2097	4	CR614786	CR614786 full-leng
C 8	252.4	28.0	1945	4	CR607080	CR607080 full-leng
C 9	250.2	27.8	1809	4	CR604283	CR604283 full-leng
C 10	247	27.4	5468	9	AQ839834	AQ839834 260L13-C5
C 11	244.8	27.2	3737	4	CR749536	CR749536 Homo sapi
C 12	244.6	27.1	669	7	CN414771	CN414771 170006008
C 13	243.6	27.0	691	6	CD102556	CD102556 AGENCOURT
C 14	243.2	27.0	710	10	AG186254	AG186254 Pan trogl
C 15	242	26.9	4640	4	AF370428	AF370428 Homo sapi
C 16	241.4	26.8	4160	4	CR859494	CR859494 Pongo pyg
C 17	240.6	26.7	655	6	CA439825	CA439825 UI-H-D10
C 18	239.2	26.5	2288	4	CR857366	CR857366 Pongo pyg
C 19	238.6	26.5	4663	4	CR857085	CR857085 Pongo pyg
C 20	238	26.4	608	10	AG161060	AG161060 Pan trogl
C 21	237.8	26.4	695	2	BG566642	BG566642 602585490
C 22	237.6	26.4	788	8	CX782424	CX782424 HESG3_14

C 23	237.4	26.3	649	11	CR960737	CR960737 Homo sapi
C 24	237	26.3	464	5	BU493340	BU493340 DKFZp781I
C 25	236.6	26.3	346	9	AQ706012	AQ706012 HS 5556 A
C 26	236.6	26.3	1114	3	BM467056	BM467056 AGENCOURT
C 27	236.2	26.2	369	5	BU536852	BU536852 AGENCOURT
C 28	236.2	26.2	539	9	AQ379787	AQ379787 RPC111-15
C 29	235.8	26.2	460	9	BU5692	BU5692 345N8.TP CI
C 30	235.8	26.2	702	6	CA428680	CA428680 UI-H-DF0-
C 31	235.8	26.2	779	9	AQ586732	AQ586732 RPC1-11-4
C 32	235.8	26.2	3355	4	CR857235	CR857235 Pongo pyg
C 33	235.6	26.1	681	9	AQ543621	AQ543621 RPC1-11-3
C 34	235.4	26.1	666	10	AG089378	AG089378 Pan trogl
C 35	235.4	26.1	753	5	BU853329	BU853329 AGENCOURT
C 36	235.4	26.1	4146	4	HSM805210	AL833304 Homo sapi
C 37	235.2	26.1	641	5	BU102838	BU102838 BX102838
C 38	235	26.1	563	7	CK902534	CK902534 im66g10.x
C 39	235	26.1	576	5	BU078278	BU078278 im66g10.y
C 40	235	26.1	744	8	CX752348	CX752348 AGENCOURT
C 41	234.6	26.0	5022	4	CR933599	CR933599 Homo sapi
C 42	234.4	26.0	687	6	CA448731	CA448731 UI-H-E10-
C 43	234.2	26.0	550	1	AV756491	AV756491 AV756491
C 44	234.2	26.0	600	9	AQ075668	AQ075668 CIT-HSP-2
C 45	234.2	26.0	4636	4	BC028737	BC028737 Homo sapi

ALIGNMENTS

RESULT 1
BU785040/c
LOCUS
DEFINITION in43g09.x1 HR85 islet Homo sapiens cDNA clone IMAGE:6125008 3',
507 bp mRNA linear EST 11-OCT-2002
mRNA sequence.
ACCSSION BU785040
VERSION BU785040.1 GI:23830576
KEYWORDS EST.
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini;
Homnidae; Homo.
REFERENCE 1 (bases 1 to 507)
AUTHORS Melton,D., Brown,J., Kenty,G., Permutt,A., Lee,C., Kaestner,K.,
Lemishka,I., Scearce,M., Brestelli,J., Gradwohl,G., Clifton,S.,
Hillier,L., Marra,M., Pape,D., Wyllie,T., Martin,J., Blistain,A.,
Schmitt,A., Theising,B., Ritter,E., Ronko,I., Bennett,J.,
Cardenas,M., Gibbons,M., McCann,R., Cole,R., Tsagarishvili,R.,
Williams,T., Jackson,Y. and Bowers,Y.
TITLE Endocrine Pancreas Consortium
JOURNAL Unpublished (2000)
COMMENT Contact: Douglas Melton, Klaus H. Kaestner, & Hiroshi Inoue
Endocrine Pancreas Consortium
Harvard University, Howard Hughes Medical Institute
Dept of Molecular and Cellular Biology, 7 Divinity Ave, Cambridge,
MA 02138
Tel: 617-495-1812
Fax: 617-495-8557
Email: dmelton@biochem.harvard.edu
Library was constructed by Dr. Hiroshi Inoue DNA sequencing by:
Washington University Genome Sequencing Center For information on
obtaining a clone please contact: Dr. Hiroshi Inoue
(hinoe@im.wustl.edu)
Seq primer: -40UP from Gibco
High quality sequence stop: 443.
FEATURES
source
Location/Qualifiers
1..507
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/clone="IMAGE:6125008"
/tissue_type="Purified pancreatic islet"
/lab_host="DH10B"
/clone_lib="HR85 islet"

```

/notes="Organ: Pancreas; Vector: pBluescript SK(-); Site_1:
NotI; Site_2: XhoI; cDNA made by oligo-dT priming.
Size-selected on agarose gel. Average insert size ~1kb. 5'
XhoI site was destroyed after directional cloning.
Amplified once. Contact information: Hiroshi Inoue, MD,
Metabolism Div. (Alan Permutt Lab), Washington University
School of Medicine, Box 8127, 660 South Euclid Ave., St.
Louis, MO 63110. E-mail: hinoue@ingate.wustl.edu. Tel:
314-362-1916 Fax: 314-747-2692."

```

[illegible]

RESULT 2	
CR623485/c	
LOCUS	2213 bp mRNA linear HTC 21-JUL-2004
DEFINITION	full-length cDNA clone CS0DF017YO23 of Fetal brain of Homo sapiens (human).
ACCESSION	CR623485
VERSION	CR623485.1 GI:50504292
KEYWORDS	HTC; CNSLT_cDNA.
SOURCE	Homo sapiens (human)
ORGANISM	Homo sapiens
REFERENCE	Mammalia; Metazoa; Chordata; Vertebrata; Euteleostomi; Eukaryota; Eutheria; Euarchontoglires; Primates; Catarrhini; Homnidae; Homo.
AUTHORS	1 (bases 1 to 2213)
TITLE	Li,W.B., Gruber,C., Jesse,J. and Polayes,D.
JOURNAL	Full-length cDNA libraries and normalization
REMARK	Unpublished
	Contact : Feng Liang Email : fliang@lifetech.com URL : http://fulllength.invitrogen.com/ InvitroGen Corporation 1600 Paraday Avenue
REFERENCE	2 (bases 1 to 2213)
AUTHORS	Genoscope.
TITLE	Direct Submission
JOURNAL	Submitted (20-JUL-2004) Genoscope - Centre National de Sequencage : Bp 191 91006 EVRY cedex - FRANCE (E-mail : segrefgenoscope.cns.fr - Web : www.genoscope.cns.fr)
COMMENT	1st strand cDNA was primed with a NotI-oligo(dT) primer. Five clones end enriched, double-strand cDNA was digested with Not I and cloned into the Not I and EcoR V sites of the pCMVSPORT 6 vector. Library was normalized. Library was constructed by Life Technologies, a division of Invitrogen. Location/Qualifiers
FEATURES	1..2213
source	1.

/organism="Homo sapiens"	
/mol_type="mRNA"	
/db_xref="taxon:9606"	
/clone="CS0DF017Y023"	
/tissue_type="Fetal brain"	
/plasmid="PCWSPORT_6"	
ORIGIN	
Query Match	31.2%; Score 281; DB 4; Length 2213;
Best Local Similarity	63.0%; Pred. No. 3e-33;
Matches 550; Conservative	0; Mismatches 285; Indels 38; Gaps 6;
QY	30 GGCACAGTGACTCAAGCCTGTAAACACACAGTACTTTGGAGATCCCAAGGTGGGTGGATCAC 89
DB	1713 GGGCACAGTGGCTCATGCTCTAACTTCAGACATTTGGGAGACCAGAAAGGGGTGCTATCA- 1655
QY	90 TTGAGGTGAGAGTTTCGAGACCAGCCTGGTCAATATGGTGAACCCCTATCTCTACTATAAA 149
DB	1654 -TGAGGTTCAGGAGTTTCGAGACCAGCTTTAGACAACTGATGAACCCCGTCTCTACTATAAA 1596
QY	150 ATCAAAAAATTAGCTCGGTGTAGTGATGCATGCTCTAGTCCAGCTACTCTCGGGAGGCTG 209
DB	1595 ATATAAAATTAGCCGGGCATGTTGGCACAATGCTCTGTAATCCAGCTACTCTCGGGAGGCTG 1536
QY	210 AGGCAAGAGAAATGCTTGAACCTGGGAGGCAGAGGTTGCAGTGAGCGAGATCCCAACCAC 269
DB	1535 AGGCAGGAGAAATCGCTTGAACTTTGGCAAGCGGAGGTTGCGGTGAGCGAGATTTGTGCCAC 1476
QY	270 TGCATCCAGCCTGGGGACACAGCGAGACTCTATCTCAAAAAATAAATAAATAAATA 329
DB	1475 TGAACCTAGCCTGGGTGACAGAGCGAGACTCTGTCTCTAAAAACAACAACAACAACA 1416
QY	330 AAGGTCGGAGAGAAACAAAACCTAATAAGATTCTCGAAGTAAGCAGAGATACGTAAAT 389
DB	1415 AAAAAAC--AAACAACCAAAACCTCTCTGCTGGAATTCGGGTATCTCTCCAAT 1358
QY	390 ATATGTAATAAAGTTTAAATGCAATTTAACTGTAATCTTATTTGTTATTTGGTATATAA 449
DB	1357 TCATCAGTGACTTCGGGTGTAAGTTTAACTTTGGCCTCAGTATCTTCTTTGTAAGAAG 1298
QY	450 AGTAAAACAGCAGAAAGTAATGCAACTTCAAACTCTACATAAATCTATATATGGAAGT 509
DB	1297 AGAAGGAGACCAAAAAAATCTA-----ATAATATGAGATTATTAGACT 1252
QY	510 GGAAGGCATCTATAATCTTACTACCCAAAGATAACAGATTACATATATCTCCAGATTTTT 569
DB	1251 CATAGATTGCTATCTATCTTAAATCTGGAAGATACTGCCAGAGATATTTTAAATGTT 1192
QY	570 GGGGCATACACTAGCTTTTTTTTATTTTGGGAAAAATTTCCATGTGCAGGCATACCTAATTTT 629
DB	1191 TGGGAAGAGNAGGAGTTAGGGGAAGGAAGAT----GGGAGGAACAGAAGGAGAAATTA 1136
QY	630 TCTAAATGTCTATAGTAGTATTCATTTAAGATGTTTCCATAATTTTAAATAATCATGCTT 689
DB	1135 AAAAGAGSCCTTCATAATATACAAGCTACCTCCTTTCC-----CTTACAT 1091
QY	690 TAAAGTAGAGNAACTAGTTGGGCATCGTGGCTCAGCCTGT--ATCCAGCAGCTTTCCGGA 748
DB	1090 CAAAAGTGGCCTCGGTGGCCAGCGCGCATGGCTCACGCTGTAAATCCAGACATTTGGGA 1031
QY	749 GGGCGAGCAATGGATCAGTTGAGGTCCGAGTTTTCGAGACCAGCCTGGGACCAATGATG 808
DB	1030 GGCTGAGAGGGGTGGATCACTGAGGTTCAGGATTCAGACCAGCCTGGCCAAACATAGTG 971
QY	809 AAACACCCCTCTCTAATAAAAAATACAAAAATTTAGCTGGGCATTTGGTGGCAAGCACTGTAGT 868
DB	970 AAACCTGTCTCTACTAAAAATACAAAAAATTTATCTGGCGCAGTGGCGCTGTGCTGTAAT 911
QY	869 CCCAGCTACTCAGGAGTCTGAGGAGGAGTATC 901
DB	910 CCCAGCTACTCAGGAGGTTGAGGAGGAAAAATC 878

QY	456	CAAGCCAAAGTAATGCAACTTCAAACTCTACATAAAATATCTATTTATGGAAGTGAAGG	515
Db	804	CAGTCAAAATCAAAAGTTTTCAGCGAGGCATAGCTGGCTCTCACTGGTCTCCCTGAGCAGA	745
QY	516	CATCTATAATCCTACTACCCCAAGATPAACCAAGTTACATATTTCTCCAGATTTTGGGGCA	575
Db	744	CAGACCTGATTCCTCCACCTGAACTGCCTACATTTCTTATCCACAGCGGCACCTGCCTCTCTGA	685
QY	576	TACACTAGCTTTTATTTTGGGAAAAATTTCCATGTGCAGGCATACCTAATTTTCTTAA	635
Db	684	GCCTTTACTAGCTCTATAAATCTTCTCTTGGCTTTCATGTAGTAGTACACTTTTAAGCAAAA	625
QY	636	TGCTATAGTAGTATTCCTAATTTAAGGATGTTCCATAAATTTTAAATAACATGCTTT	690
Db	624	GGTAAATATCCCTTAAACAGTATGCATATAAACCATAATTTTGTGTAGTAGCTGTTTCT	565
QY	691	-----AAAGTAGAGAACTAGGTTGGGCATGGTGGCTCAGGCTGT-ATCCCCAGCATTT	743
Db	564	TTCCCATTAACCAAAATGTTGGCCGGCAGCGTGTTCATGCTGTAAATCCAGCACCT	505
QY	744	TGGGAGGCCGAGGCAAAATGGATCACTTGAGTTCGGAGTTTGAGACAGCAGCTCGACAACA	803
Db	504	TGGGAGGAGGAAGTTGGGGGATCACCTGAGTTCAGGATTCGAGACAGCAGCTCGACAACA	445
QY	804	TGATGAAACACCCCTCTCTAATAAAATAC-AAAAATTAGCTGGGCATGTGGCAAGCACC	862
Db	444	CGGCAAAACCTTGTCTCTACCAAAAATACAAAAAATTTAGCTGGGCATGTGGCGCC	385
QY	863	TGTAGTCCCACTACTCAGGAGTCTGAGGCAGGAGTAT	900
Db	384	TGTAATCCAGCTACTTTGGGAGGCTGAGGCAGGAGAT	347
RESULT 4			
CR627178/c			
LOCUS	CR627178	6056 bp	mRNA linear HTC 22-SEP-2004
DEFINITION	Homo sapiens mRNA; cDNA DKFZp686B03112 (from clone DKFZp686B03112).		
ACCESSION	CR627178		
VERSION	CR627178.1	GI:50949761	
KEYWORDS	HTC.		
SOURCE	Homo sapiens (human)		
ORGANISM	Homo sapiens		
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini; Hominidae; Homo.			
REFERENCE	1 (bases 1 to 6056)		
AUTHORS	Poustka,A., Albert,R., Moosmayer,P., Schupp,I., Wellenreuther,R., Mewes,H.W., Well,B., Amid,C., Osanger,A., Fobo,G., Han,M. and Wiemann,S.		
CONSTRM	The German cDNA Consortium		
TITLE	Direct Submission		
JOURNAL	Submitted (22-SEP-2004) MIPS, Ingolstaedter Landstr.1, D-85764 Neuherberg, GERMANY		
COMMENT	Clone from S. Wiemann, Molecular Genome Analysis, German Cancer Research Center (DKFZ); Email s.wiemann@kfz-heidelberg.de; Sequenced by DKFZ (German Cancer Research Center, Heidelberg/Germany) within the cDNA sequencing consortium of the German Genome Project. This clone (DKFZp686B03112) is available at the RZPD Deutsches Ressourcenzentrum fuer Genomforschung GmbH in Berlin, Germany. Please contact RZPD for ordering: http://www.rzpd.de/cgi-bin/products/cl.cgi?CloneID=DKFZp686B03112 Further information about the clone and the sequencing project is available at http://mips.gsf.de/projects/cdna/.		
FEATURES	Location/Qualifiers		
source	1. .6056		
	/organism="Homo sapiens"		
	/mol_type="mRNA"		
	/db_xref="taxon:9606"		
	/clone="DKFZp686B03112"		
	/tissue_type="fetal kidney"		
	/clone_lib="686 (synonym: hlcc3). Vector pSport1_Sfi; host DH10B; sites SfiIA + SfiIB"		
	/dev_stage="fetal"		

ORIGIN	/note="3' UTR"									
	BX431110/c									
	LOCUS									
	DEFINITION									
	BX431110 Homo sapiens FETAL BRAIN mRNA linear EST 05-MAY-2004									
	BX431110 CS0DF017Y023 5-PRIME, mRNA sequence.									
	ACCESSION									
	VERSION									
	KEYWORDS									
	SOURCE									
ORIGIN	ORGANISM									
	Homo sapiens (human)									
	Homo sapiens									
	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini; Hominidae; Homo.									
	REFERENCE									
	AUTHORS									
	TITLE									
	JOURNAL									
	COMMENT									
	On May 22, 2003 this sequence version replaced gi:31018008.									
FEATURES	Genoscope - Centre National de Sequencage									
	2 rue Gaston Cremieux, CP 5706 - 91057 EVRY cedex - FRANCE									
	Email: seqref@genoscope.cns.fr, Web : www.genoscope.cns.fr									
	1st strand cDNA was primed with a NotI-oligo(dT) primer. Five prime end enriched, double-strand cDNA was digested with Not I and cloned into the Not I and EcoRV sites of the pCMVSPORT 6 vector. Library was not normalized. Library was constructed by Life Technologies, a division of Invitrogen.									
	This sequence belongs to sequence cluster 8501.r									
	For more information about this cluster, see									
	http://www.genoscope.cns.fr/cdna?s=CS0BAF0142E10_AF01326_2&c=8501.r									
	Location/Qualifiers									
	1..903									
	/organism="Homo sapiens"									
ORIGIN	/mol_type="mRNA"									
	/db_xref="taxon:9606"									
	/cson="CS0DF017Y023"									
	/tissue_type="FETAL BRAIN"									
	/dev_stage="fetal"									
	/clone_lib="Homo sapiens FETAL BRAIN"									
	/note="Organ: brain; Vector: pCMVSPORT_6; 1st strand cDNA was primed with a NotI-oligo(dT) primer. Five prime end enriched, double-strand cDNA was digested with Not I and cloned into the Not I and EcoRV sites of the pCMVSPORT 6 vector. Library was not normalized."									
	Query Match									
	Best Local Similarity									
	Matches									
ORIGIN	543; Conservative									
	0; Mismatches									
	292; Indels									
	38; Gaps									
	7;									
	30 GGGCACAGTGAAGTTCGAGACAGCGCTGCTCAATATGTTGAAACCCCTATCTCTACTATAA									
	89									
	903 GGGCACAGTGGCTCATGCTGTAAATCTCAGCACATTGG-AGACCAACGGGTGATCA-									
	846									
	90 TTGAGTTCAGAAAGTTCGAGACAGCGCTGCTCAATATGTTGAAACCCCTATCTCTACTATAA									
ORIGIN	149									
	845 -TGAGGTTCAGAGTTCGAGACAGCGCCAGACCAATGATGAACCCCGTCTCTACTATAA									
	787									
	150 ATACAAAAATAGCTGGGTGTAGTGAATGATGCTGCTGTAGTCCAGCTACTCGGGAGGCTG									
	209									
	786 ACACAAAAATTAGCCCGGCGCATGGTGGACATGCTGTAAATCCAGCTACTCGNAGGCTG									
	727									
	210 AGGCAAGAGAAATGCTTGAACCTGGAGGCGAGAGTTCGAGTGCAGGATGCCAGATCCACAC									
	269									
	726 AGGCAGGAGAAATCGCTTGAACCTGGAAAGCGGAGGTTCGGGTGAGCCGAGATTTGGCCAC									
ORIGIN	667									
	270 TGCACCTCCAGCTCGGCGCACAGCGAGACTCTATCTCAAAAAATAATAATAATAATA									
	329									
	666 TGAACCTCCAGCTCGGCGCACAGCGAGACTCTATCTCAAAAAATAATAATAATAATA									
	608									
	330 AAGGATCGGAGAGAAACAAACTAATAAGATTCTTGAAGGTGAAGCAGATACGTAAT									
	389									
	607 ACACAAAAACAAACAAACAAACAAACCTCTCTGCTGGAATTCGGGTATCTCTCCTCACT									
	548									
	548									


```
RESULT 7
CR614786/c
LOCUS
DEFINITION
full-length cDNA clone CS0DF006YD23 of Fetal brain of Homo sapiens
(human).
ACCESSION
CR614786.1 GI:50495593
VERSION
HTC; CNSLT cDNA.
KEYWORDS
Homo sapiens (human)
SOURCE
Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini;
Hominidae; Homo.
REFERENCE
1 (bases 1 to 2097)
AUTHORS
Li, W.B., Gruber, C., Jesse, J. and Polayes, D.
TITLE
Full-length cDNA libraries and normalization
JOURNAL
Unpublished
REMARK
Contact: Feng Liang Email: fliang@lifetech.com URL:
http://fulllength.invitrogen.com/ Invitrogen Corporation 1600
Paraday Avenue
2 (bases 1 to 2097)
Genoscope.
Direct Submission
Submitted (20-JUL-2004) Genoscope - Centre National de Sequencage :
BP 191 91006 EVRY cedex - FRANCE (E-mail: segref@genoscope.cns.fr
- Web: www.genoscope.cns.fr)
1st strand cDNA was primed with a NotI-oligo(dT) primer. Five prime
end enriched, double-strand cDNA was digested with Not I and cloned
into the Not I and EcoR V sites of the pCMVSPORT 6 vector. Library
was normalized. Library was constructed by Life Technologies, a
division of Invitrogen.
FEATURES
Location/Qualifiers
1..2097
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/clone="CS0DF006YD23"
/tissue_type="Fetal brain"
/plasmid="pCMVSPORT_6"
ORIGIN
Query Match 28.1%; Score 253; DB 4; Length 2097;
Best Local Similarity 61.1%; Pred. No. 5.3e-29;
Matches 542; Conservative 0; Mismatches 290; Indels 55; Gaps 6;
QY 16 AATAAATCACTGTTGGGCACAGTGAAGTTCGAGACAGAGCTTGTGAAGTCCAA 75
DB 1329 AATCAACATGGCTGGGCATGGTGTTCACGTTGTATCCAGCACTTTGGGAGCTGA 1270
QY 76 GTGGGTGGATCACTTGGAGTGAGAGTTCGAGACAGAGCTTGTCAATATGGTGAACCC 135
DB 1269 GTGGGCAGATCAGTTGAGTTCAGGAGTTCAAGACCAGCTGGCCAAATGATGAACCC 1210
QY 136 TATCTCTCTAAATAATACAAAATTAAGTGGTGTAGTGATGATGCTGTGTAGTCCAGC 195
DB 1209 CGTCTCTCTAAATAATACAAAATTAAGTGGTGTGGTGTGGTGGTGGTGGTGGTGGT 1150
QY 196 TACTCGGGAGGC-TGAGGCAAGAGATTGCTTGAACCTGGAGGAGAGTTGCAGTGAG 254
DB 1149 TACTTGGGAGGCTGAGGCTGAGGAGTGAAGATCACTTGAACCTGGAGGAGAGTTGCAGTGAG 1090
QY 255 CCAGATCCCACTGCACTCCAGCTGGGCGACACAGCGAGCTCTATCTCAAAAAAA 314
DB 1089 CCAAGATCAGCCACTGCTTCCAGCTGGCAACAGAGCAAGACTCCATCTC-AAAAA 1031
QY 315 TAAATAAATAAATAAAGATCGGAGAGAAACAAAACCTAATAGATCTCTGAAGTAAAGC 374
DB 1030 GAAATAAATAAAGCCAGGCGCAGTGGCTCAGCCCTGTGTAATCCAGCACT-TTGGGAGGC 972
QY 375 AGAGATACGTAATATATATATATAAGTTTAAATGATTAATGATTAATGATTAATGTT 434
```

```
971 CGGGTGGGTGAATCACCTGAGGTGAGCGGTTCAAGACCAGCCTGTCACCATGGTGA 912
435 TATTTTGGTTATAAAGTAAACAGCCAAAGTAATGCAACTTCAAACTCTACATAATA 494
911 CCCTGTCTCTACTACAC-----AATACAAAAA 886
495 TCTATTATGAAAAGTGAAGGCATCTATATCTTACCCAAAGATAAACAGTTACATA 554
885 TTAGTGGGCTGTGTGGTGGTCCCTCTATCCCACTCTCAGGAGGCTGAGCAGGAG 826
555 TTCTCCAGATTTTGGGCGATACACTAGCTTTTATTTTGGGAAAATTTCCATGTGCA 614
825 AATCGCTTGAATCCAGGAG-----TGAGAGGTAGCAGTGCCTCAAGAT 783
615 GGCATACCTAATTTTCTAATCTATGTAGTATTCCATTTAAGGATGTTCCATAATTT 674
782 TGCACCACTGCTCCAGGCTGGAACAACAGAGCGAGACTCCGTCTCAAAAAAAGAGAA 723
675 TTAATAATACATGCTTTAAAGTAGAGAAACTAGTGTGGCATGGTGGCTCAGCCTGT-AT 733
722 AATATTTAAGAAGTTTAAAGAGAGAGAGACCTCCGGGCACTGGTGGCTCATACCTTAAT 663
734 CCCAGCACTTTGGGAGCCCGAGCAAAATGGATCACTTTGAGTCCGGAGTTTGACACAGC 793
662 CCCAGCGCTTCGGGAGGAGGAGGCAAGGCGGTGGATCATTTGAGGTTCAGGAGTTCAAGACCAGC 603
794 CTGACACATGATGAACACCTCTCTAATAAATAACAAAATTAAGTGGGATGTTG 853
602 CTGACCAACATGTTGAAACTCCGCTCTATTTAAATAATACAAAATTAAGCAGGATGTTG 543
854 GCAAGCACCTGTAGTCCAGCTACTCAGGAGTCTGAGGCGAGGAGTAT 900
542 GCATGGCGCTATATCCAGCTACTCAGGAAGTCTGAGGCGAGGAGAT 496

RESULT 8
CR607080
LOCUS
DEFINITION
full-length cDNA clone CS0DI007YG08 of Placenta Cot 25-normalized
of Homo sapiens (human).
ACCESSION
CR607080.1 GI:50487887
VERSION
HTC; CNSLT cDNA.
KEYWORDS
Homo sapiens (human)
SOURCE
Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini;
Hominidae; Homo.
REFERENCE
1 (bases 1 to 1945)
AUTHORS
Li, W.B., Gruber, C., Jesse, J. and Polayes, D.
TITLE
Full-length cDNA libraries and normalization
JOURNAL
Unpublished
REMARK
Contact: Feng Liang Email: fliang@lifetech.com URL:
http://fulllength.invitrogen.com/ Invitrogen Corporation 1600
Paraday Avenue
2 (bases 1 to 1945)
Genoscope.
Direct Submission
Submitted (20-JUL-2004) Genoscope - Centre National de Sequencage :
BP 191 91006 EVRY cedex - FRANCE (E-mail: segref@genoscope.cns.fr
- Web: www.genoscope.cns.fr)
1st strand cDNA was primed with a NotI-oligo(dT) primer. Five prime
end enriched, double-strand cDNA was digested with Not I and cloned
into the Not I and EcoR V sites of the pCMVSPORT 6 vector. Library
was normalized. Library was constructed by Life Technologies, a
division of Invitrogen.
FEATURES
Location/Qualifiers
1..1945
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/clone="CS0DI007YG08"
/tissue_type="Placenta Cot 25-normalized"
```


RESULT 9					
CR604283/c					
LOCUS	CR604283	1809 bp	mRNA	linear	HTC 21-JUL-2004
DEFINITION	full-length cDNA clone CSODF012YB15 of Fetal brain of Homo sapiens (human) .				


```
Db 472 GCCAGAGAGAAATAAGCAAAATGAATGAATACTAGAGCAAAACAAACAGAGAAACCA 413
Qy 558 -CTCAGATTTTGGGGCATACACTAGCTTTTATTTGGGAAATTTCCATGTGAGG 616
Db 412 TCTCTACATTTTCTAATCGAAGCTGAATGCTACAGCAATATCCAGGGCCGAGA 353
Qy 617 CATACCTAAATTTTCTAAATGCTATAGTATTCATTTAAGGATGTTCCATAATTTT 676
Db 352 AAACACACATAATACTTTATTTCTGATAAACAAATTCGATGCTTAAAGTCGCCAATATC 293
Qy 677 AAAATACATGCTTTAAAGTAGAGAA- - -ACTAGCTTGGGCATGCTGCTCACGCCCTGT- 731
Db 292 ACCGTACATCTGTGCAATCAAGAAATTTATCTAGCTGGGCACAGTGGCTCACACCTGTA 233
Qy 732 ATCCAGACACTTTGGGAGCCGAGCAAAATGGATCATTGAGGTCGGGAGTTTGAGACCA 791
Db 232 ATTCAGACACTTTGGGAGCCGAGGCGGTGGAATCACCTGAGTCAGGTCAGGCAATTCGAGAA 173
Qy 792 GCCTGGACAACATGATGAACACCCCTCTCTAATAAAAAATACAAAAATTTAGCTGGGCATGG 851
Db 172 GCCTGGCCACATGCGGAACCCCATCTCTACTAAAAATACAAAAATTTAGCCGGTATGG 113
Qy 852 TGGCAAGACCTGTAGTCCAGCTACTCAGGAGTCTGAGGCGAGGATAT 900
Db 112 AGGCATGTGCTGTAATCCAGCTACTTGGGAGGCTGAGGCGAGGAGAAAT 64
```

```
RESULT 10
A0839834/c
LOCUS
DEFINITION
260L13-C52 C17B Homo sapiens genomic clone 260L13, genomic survey
sequence.
ACCESSION
A0839834
VERSION
A0839834.1 GI:6652466
KEYWORDS
GSS.
SOURCE
Homo sapiens (human)
ORGANISM
Homo sapiens
REFERENCE
1 (bases 1 to 5468)
Cartpen,J.D., Makalowska,I., Robbins,C.M., Scott,N., Sood,R.,
Connors,T.D., Bonner,T.I., Smith,J.R., Faruque,M.U., Stephan,D.A.,
Pinkett,H., Morgenbesser,S.D., Su,K., Graham,C., Gregory,S.G.,
Williams,H., McDonald,L., Baxevaris,A.D., Klingler,K.W. and
Landes G.M.
A 6-Mb high-resolution physical and transcripition map encompassing
the hereditary prostate cancer 1 (HPC1) region
Genomics 64 (1), 1-14 (2000)
10708513
Contact: Cartpen JD
Cancer Genetics Branch
National Human Genome Research Institute/National Institutes of
Health
Bldg. 36, Room 3D04, 36 Convent Drive, Bethesda, MD
Tel: 301 435 5626
Fax: 301 435 5465
Email: jdc@nhgri.nih.gov
Class: shotgun.
Location/Qualifiers
1. 5468
/organism="Homo sapiens"
/mol_type="genomic DNA"
/db_xref="taxon:9606"
/clone="260L13"
/clone_lib="C17B"
```

FEATURES

source

1. 5468
/organism="Homo sapiens"
/mol_type="genomic DNA"
/db_xref="taxon:9606"
/clone="260L13"
/clone_lib="C17B"

ORIGIN

Query Match 27.4%; Score 247; DB 9; Length 5468;
Best Local Similarity 57.4%; Pred. No. 3e-28;
Matches 514; Conservative 0; Mismatches 360; Indels 21; Gaps 3;

```
Qy 10 TAAATAAAAATAACTCAGTTGGGCACAGTCACTCAAGCCCTGTAAACACAGTACTTTGGAA 69
Db 2985 TAAAGGACAAGTAAGGGCGCGGTGGCTCACGCCCTGTAAATCCAGCACTTTGGGA 2926
Qy 70 GTCACAGGTGGGTGATCACTTGAAGTGAAGTTGAGACACAGCCTGGTCAATAT-GGT 128
Db 2925 GGCCAAAGGCCAGCGGATCACCTGAGGTTGAGGTTGAGACTAGCCCGGCAACATGGGT 2866
Qy 129 GAAACCTATCTCTACTAAAAATACAAAAATAGCTGGGTGATGATGATGCTCTGTAG 188
Db 2865 GAAACCTCTGTTCTTACTAAAAATACAGAAATAGCCAGGTGATGTTGTTGCTCTAG 2806
Qy 189 TCCAGCTACTCGGGAGGCTGAGGCAAGAGAAATGCTTGAACCTGGGAGGAGAGGTTGC 248
Db 2805 TCCAGCTGCTTGGAGACTGAGGCTGGAGAAATCGCTTGAACCTGGGAGGAGAGGTTGC 2746
Qy 249 AGTGAGCCGAGATCCCAACCACTGCACTTCAGCCCTGGGCGACACAGCAGACTCTATCTCA 308
Db 2745 AGTGAGCCGAAATCTTGTCTCATTCGATTTCCAGCCCTGGGTGACAAAGTGAGACTCCATCATA 2686
Qy 309 AAAAAATAAATAAATAAATAAAGGATCGGAGAGAAACAAAACCTAATAAGATTCTCTGAAG 368
Db 2685 AATCAGTCNAATCATGTAGTTTCAAGTTTGAATTTCTCTTCAACCCATCTTAAATGTCCTT 2626
Qy 369 GTAAGCAGAGATACG--TAAATATATATGTAATAAGTTTAAATGCAATTTTAACTGTAATC 426
Db 2625 TTGCCCCCAATTTAGTCTTACTTGAATCTTATGAAGATTACCATCTTTGGCTTCTCTTTAG 2566
Qy 427 TTATTGTTTATTTGGTTATAAAGTAAACAGCCAAAGTAAGTAAGCACTTCAAACTCTA 486
Db 2565 CTCTGTAACTCTTCTTGTGAATGTTCCCTTTTATTTACTAGCCACACTGACCTGCTTTC 2506
Qy 487 CATAAATATCTTATATGGAAGTGAAGGCATCTATAATCTCTACTACCCAAAGATAACCA 546
Db 2505 TTTCCCTTCTCTAATATATGTTTGAAGTTATCTCTGTTTCCCTATCTTACATGCGC 2446
Qy 547 GTTACATATTCCTCCAGATTTTGGGGCATACACTAGCTTTTATTTTGGGAAAAATTC 606
Db 2445 TCCTCTTCTCCCTTGAATGGAAGTCCCTAGTTTAACTCAT----- 2404
Qy 607 CATGTGAGGCATACCTTAAATTTTCTAAATGCTATGATAGTATTCATTTTAAAGGATGTT 666
Db 2403 AAGATATCAGCATTTAGTAAGCAGTAATTTCTTCTGAGTGATGATCTTAGTGATCTTCAAT 2344
Qy 667 CATAAATTTTAAATACATGCTTTTAAAGTACAGAAACTAGGTGGGCATGTTGCTCAGC 726
Db 2343 ATGAATGAAGGTGGCTGGCGCATCTCCCTCCACTATTCAGCTGGGATTT 2284
Qy 727 CCGTATCCAGCACTTTGGGAGGCGGAGGCAAAATGGATCACTTGAGTCCGAGTTTGA 786
Db 2283 TACAATCCAGCACTTTGGGAGGCGGAGGTCAGGTCAGGTTGGGAGTTGCA 2224
Qy 787 GACAGCCTGGACAACATGATGAACACCCCTCTCTAATAAAAAATACAAAAATTTAGCTGG 846
Db 2223 GACAGCCTGGCAACATGTTGAACCCCATCTCTACTAAAAATAGAAAAATTCGCCAGG 2164
Qy 847 CATGTGGCAGCACTCTAGTCCAGCTACTCAGAGTCTCGAGCAGGAGTATC 901
Db 2163 CGTAGTGGCTCACGCTGTAAATCCAGCACTTTGGGAGGCCAAGCAGCGGATC 2109
```

RESULT 11

CR749536/c

LOCUS

DEFINITION

ACCESSION

VERSION

KEYWORDS

SOURCE

ORGANISM

Homo sapiens

Homo sapiens

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini;

Hominoidea; Homo.

CR749536 3737 bp mRNA linear HTC 19-0004

Homo sapiens mRNA; cDNA DKFP686L24211 (from clone DKFP686L24211).

CR749536

CR749536.1 GI:51476749

HTC.

Homo sapiens (human)

Homo sapiens

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini;

Hominoidea; Homo.

REFERENCE 1 (bases 1 to 3737)
 AUTHORS Koehler,K., Beyer,A., Mewes,H.W., Weil,B., Amid,C., Osanger,A.,
 Fobo,G., Han,M., and Wiemann,S.
 CONSRTM The German cDNA Consortium
 TITLE Direct Submission
 JOURNAL Submitted (17-AUG-2004) MIPS, Ingolstaedter Landstr.1, D-85764
 Neuherberg, GERMANY
 COMMENT Clone from S. Wiemann, Molecular Genome Analysis, German Cancer
 Research Center (DKFZ); Email s.wiemann@dkfz-heidelberg.de;
 sequenced by BMFZ (Biomedical Research Center at the
 Heinrich-Heine-University, Dueseldorf/Germany) within the cDNA
 sequencing consortium of the German Genome Project. This clone
 (DKFZp686L2421) is available at the RZPD Deutsches
 Ressourcenzentrum fuer Genomforschung GmbH in Berlin, Germany.
 Please contact RZPD for ordering:
 http://www.rzpd.de/cgi-bin/products/cl.cgi?cloneID=DKFZp686L24211
 Further information about the clone and the sequencing project is
 available at http://mips.gsf.de/projects/cdna/
 FEATURES
 Location/Qualifiers
 1..3737
 /organism="Homo sapiens"
 /mol_type="mRNA"
 /db_xref="taxon:9606"
 /clone="DKFZp686L24211"
 /tissue_type="rectum tumor"
 /clone_lib="686 (synonym: hlccc3). Vector pSport1_Sfi; host
 DH10B; sites SfiIA + SfiIB"
 /dev_stage="adult"
 /note="transmembrane protein dectin-1, differentially
 spliced, not fully spliced"
 1..3737
 /gene="DKFZp686L24211"
 /cds_start=1
 /gene="DKFZp686L24211"
 /codon_start=1
 /product="hypothetical protein"
 /protein_id="CAH18342.1"
 /db_xref="GI:51476750"
 /translation="MEYHPDLENLDEDCYTOLHFDQSNTRIAVVSEKSGCAASPMPWR
 LIAVILGLICLVILVIAVLGTMGELVLEEINRQFSKQLLCARHSFKGLTCLIFNLI
 YLIWFLYTWGAVNHNFTYKGCERLGRYTFPGL"
 ORIGIN
 Query Match 27.2%; Score 244.8; DB 4; Length 3737;
 Best Local Similarity 60.6%; Pred. No. 7, 4e-28;
 Matches 544; Conservative 0; Mismatches 337; Indels 17; Gaps 8;
 QY 4 TCATATTAAATTAATCACTCAGTTGGGCGACAGTCACTCAAGCTGTAAACCAAGTACT 63
 DB 3229 TTACATATTATACAAAGATTCTAGGCTGGGTGCGTCAAGCTGTAAATCCAGCAC- 3171
 QY 64 TTGGAAGTCCAGGTGGGTGATCACTTGGTGGAGAGTTGGAGACCGCTGTGTCAT 123
 DB 3170 TTGGAGGCGCGAGGCGAGGAGATCACTTGGTGGAGAGTTGGAGTCCGAGTCA 3111
 QY 124 ATGGTGAACCTCTCTCTCTAAATAACA-AAAATTAGTGGTGTAGTGATGCATGC 182
 DB 3110 ATGGAGAACCTGTCTCTACTTAAATAATAAATAATAGCGGTGGTGGTGGCTGA 3051
 QY 183 CTGTAGTCCAGTACTCTGGAGGCTGAGGCAAGAGATTCTTGAACCTGGGAGGCGA 242
 DB 3050 CTGTAATCCAGTACTTGGAGGCTGAGGCGAGGAGATTCTTGAACCCAGGCGCGA 2991
 QY 243 GGTTCAGTGGCGGAGATCCACCACTGCATTCAGCTCCGCTGGGCGAC-ACAGCGAGATC 301
 DB 2990 GGTTCGCTGAGCGAGACCGTGGCGGTGGCTCAGCTCCGCGCAACAGAGCGAAATC 2931
 QY 302 TATCTCAAAAAATAAATAAATAAATAAAGGATCGGAGAGAAACAAACTTAATAAGATT 361
 DB 2930 TGTCTCAAAAA-AAAAAATAAATAAATAAAGATTCTATTAGATATGGCAGATGTACT 2877
 QY 362 CCTGAAGTAAAGCAGAGATACGTAATATTATATGTAATAAAGTTTAAATGCACTTTAACTG 421

DB 2876 TCTGAGCTATTAAATTGATAAAGAAAAGTAGGAGTCTCTGCTTTTACTAAGATGATAA 2817
 QY 422 TAATCTTATTGTTTATTTTGGTTATATAAGTAAACCAAGCCAAAGTAACTCACTTCAA 481
 DB 2816 TCATAATTATGATCATACTGATGAAATAATAAATAATTGCTCTGACTGTGGTGAA--GATT 2759
 QY 482 CTCTACATAAATATCTATTATGGAAGTGGAGGCATCTATAATCTCTACTACCCCAAGAT 541
 DB 2758 TTTTCCCATTTCTTCATTAGCTATTAGAAATATATTTACCCCTCAGTTCTATAACTGAT 2699
 QY 542 AACAGTTACATATCTCCAGATTTTGGGCGCATACACTAGCTTTTATTTTGGGAAA 601
 DB 2698 ATACTGCTAGAAAGTTGAGGGTCAAAATCGTGGCAACACACCCGTCACCTTCAAT---GGCA 2643
 QY 602 ATTTCCATGTGCGAGCATACCTAAATTTTCTAAATGTCTATGTAGTAGTATTTCCATTTAAGGA 661
 DB 2642 TTGTTGAGCATGAACAACTCTGATAGAGCCAGAAATTAAGATTGCGCTTATGGCTTATCCC 2583
 QY 662 TGTTCCATAAATTTTAAATAACATGCTTTTAAAGTAGAGAAACTAGGTTGGGCATGGTGGC 721
 DB 2582 AAGGTAAATGCCAAGTCAATGGCAGAAATGAAATTTCAAGCTTGGGCGCAGGTGTGGTGGC 2523
 QY 722 TCAGCCCTGT-ATCCAGCACTTTGGAGGCGCGAGGCAAAATGATCACTTGAAGTCCGGA 780
 DB 2522 TCACATCTGTAATCCAGCACTTTGGGAGGCGCAAGGCGGCGAGATCACTTAAGCCCGAGA 2463
 QY 781 GTTTGAGACAGCCTGCGACACATGATGAACACCCCTCTCTAATAAATAA-CAAAAATT 839
 DB 2462 GTTAGAGACCAAGCTTGGTCAACATGGCAAAACCCAGTCTCTACAAAAAACACCAAAAATT 2403
 QY 840 AGCTGGGCGTGGTGGCAAGCACCTGTAGTCCAGCTACTCAGGAGTCTTGAGGCGAGGAG 897
 DB 2402 AGCCTGGCAGTGGTGGCATGCACCTGTAGTCCAGCTACTTTCAGGGGCTGAGGCGGAG 2345

RESULT 12
 CN414771 LOCUS 669 bp mRNA linear EST 16-MAY-2004
 DEFINITION 17000600839160 GRN_PREHEP Homo sapiens cDNA 5', mRNA sequence.
 ACCESSION CN414771
 VERSION CN414771.1 GI:47402365
 KEYWORDS EST.
 SOURCE Homo sapiens (human)
 ORGANISM Homo sapiens
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini;
 Homnidae; Homo.
 REFERENCE 1 (bases 1 to 669)
 AUTHORS Brandenberger,R., Wei,H., Zhang,S., Lei,S., Murage,J., Fisk,G.J.,
 Li,Y., Xu,C., Fang,R., Guegler,K., Rao,M.S., Mandalam,R.,
 Lebkowski,J and Stanton,L.W.
 TITLE Transcriptome characterization elucidates signaling networks that
 control human ES cell growth and differentiation
 JOURNAL Nat. Biotechnol. 22 (6), 707-716 (2004)
 PUBMED 15146197
 COMMENT Contact: Brandenberger R
 Regenerative Medicine
 Genon Corporation
 230 Constitution Drive, Menlo Park, CA 94025, USA
 Tel: 650 473 8658
 Fax: 650 473 7760
 Email: rbrandenberger@genon.com
 Insert Length: 669 Std Error: 0.00.
 Location/Qualifiers
 1..669
 /organism="Homo sapiens"
 /mol_type="mRNA"
 /db_xref="taxon:9606"
 /tissue_type="embryonic stem cells, DMSO-treated H9 cell
 line"
 /clone_lib="GRN_PREHEP"
 /notes="oligo dT primed, full-length enriched cDNA library
 from DMSO-treated hES cell line H9 (p22) maintained in

feeder-free conditions"

```

ORIGIN
Query Match      27.1%; Score 244.6; DB 7; Length 669;
Best Local Similarity 83.7%; Pred. No. 1.5e-27;
Matches 277; Conservative 0; Mismatches 54; Indels 0; Gaps 0;

Qy 1 TGTTCATATTAATTAATAAATTAATCAGTTGGGACAGTCAAGCTGTAACACAGT 60
Db 46 TGATCATCTCTGTGAAATTTGCAACAGCGGGACAGTGGCTGCTGTAATCCCGAGC 105

Qy 61 ACTTTGGAGTCCAGGTGGGTGGATCACTTGAGTGAGAAAGTTGAGACACAGCTGTGC 120
Db 106 ACTTTGGAGTCCAGGTGGGTGGATCACTTGAGTGAGAAAGTTGAGACACAGCTGTGC 165

Qy 121 AATATGGTGAACCCCTATCTCTACTTAAATAATACAAATAATAGCTGGGTGTAGTGATGAT 180
Db 166 AACATGGTGAACCCCTGTCTCTACTTAAATAATACAAATAATAGCTGGGTGTAGTGATGAT 225

Qy 181 GCCTGTAGTCCAGTACTCTGGGAGGCTGAGGCAAGAGAAATTTGCTTGAACCTGGGAGGCA 240
Db 226 GCCTGTAGTCCAGTACTCTGGGAGGCTGAGGCAAGAGAAATTTGCTTGAACCTGGGAGGCA 285

Qy 241 GAGGTTCAGTGAGCGGAGATCCACCACTGCACTCCAGCTGGGCGACACAGCGAGACT 300
Db 286 GAGGTTCAGTGAGCGGAGATCCACCACTGCACTCCAGCTGGGCGATGACAGGCAAGACT 345

Qy 301 CTATCTCAAAAAATAAATAAATAAATAAATAA 331
Db 346 CTGCTCAATAATAATAATAATAATAAATAA 376

```

```

RESULT 13
LOCUS CD102556/c
DEFINITION AGENCOURT 14024251 NIH_MGC_186 Homo sapiens cDNA clone
IMAGE:30369934 5', mRNA sequence.
CD102556
CD102556.1 GI:30755730
EST.
Homo sapiens (human)
Homo sapiens

```

```

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini;
Hominoidea; Homo.
REFERENCE 1 (bases 1 to 691)
AUTHORS NIH-MGC http://mgc.nci.nih.gov/.
TITLE National Institutes of Health, Mammalian Gene Collection (MGC)
JOURNAL Unpublished (1999)
COMMENT Contact: Robert Strausberg, Ph.D.
Email: cgabs-r@mail.nih.gov
Tissue Procurement: Dr. Michael Brownstein and Dr. Miklos Palkovits
CDNA Library Arrayed by: The I.M.A.G.E. Consortium, Inc.
CDNA sequencing by: Agencourt Bioscience Corporation
Clone distribution: MGC clone distribution information can be
found through the I.M.A.G.E. Consortium/LLNL at:
http://image.llnl.gov
Plate: NDQM158 row: b column: 23
High quality sequence stop: 509.

```

```

FEATURES
source
Location/Qualifiers
1..691
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/clone="IMAGE:30369934"
/lab_host="DHIOB (T1 phage-resistant)"
/clone_lib="NIH_MGC_186"
/notes="Organ: Pooled-Skin; Vector: pDNR-LIB; Site 1: Sfil
(ggcccattggcc); Site 2: Sfil (ggccgctcggcc); Library is
oligo-dt primed and directionally cloned. cDNA was
prepared from a pooled samples of tissues from Skin,
meninges, duramatter, pia matter and choroid plexus. 5'

```

and 3' adaptors were used in cloning as follows: 5' adaptor sequence: 5'-CAGGCCATTATGGCC-3' and 3' adaptor sequence: 5'-ATTCTAGAGCCGAGCGCGCATG-dt(30)BN-3' (where B = A, C, or G and N = A, C, G, or T). Average insert size 1.47 kb (range 0.50-4.0 kb). 15/15 colonies contained inserts by PCR. This library was enriched for full-length clones and was constructed by Clontech Laboratories (Palo Alto, CA). Note: this is a NIH_MGC Library"

```

ORIGIN
Query Match      27.0%; Score 243.6; DB 6; Length 691;
Best Local Similarity 82.5%; Pred. No. 2.1e-27;
Matches 279; Conservative 0; Mismatches 59; Indels 0; Gaps 0;

Qy 5 CATATTAATTAATAAATAAATCACTAGTTGGGACAGTCAAGCTGTAAACACAGTACTT 64
Db 661 CAAATAATAAGGAACATCTCAGCTGGGCAAGTGGCTCAAGCTGTAAATCCAGCACTT 602

Qy 65 TGAAGTCCAGGTGGGTGGATCACTTGAGGTGAGAAGTTCCGAGACCAAGCTGGTCAATA 124
Db 601 TGGAGGCGAAGGTGGGAGATCACTTGAGGCCAGGAGTTCAAGCCAGCTTAGCCACA 542

Qy 125 TGGTGAACCCCTATCTCTACTTAAATAATACAAATAATAGCTGGGTGTAGTGATGCT 184
Db 541 TGGTGAATCCATCTCTACTTAAATAATACAAATAATAGCTGGGTGTAGTGATGCT 482

Qy 185 GTAGTCCAGTACTCTGGGAGGCTGAGCAAGAGAAATTTGCTTGAACCTGGGAGGAGG 244
Db 481 GTAGTCCAGTACTCTGGGAGGCTGAGCAAGAGAAATTTGCTTGAACCTGGGAGGAGG 422

Qy 245 TTGCAGTGAGCGGAGATCCACCACTGCACTCCAGCTGGGCGACACAGCGAGACTTAT 304
Db 421 TTGCAGTGAGCGGAGATCGTCCACTGCACTCCAGCTGGGCGATGAGGAGACTCTGT 362

Qy 305 CTCAAAAATAAATAAATAAATAAATAAAGATCGGAGAG 342
Db 361 CTCATTGAATAAATAAATAAATAAATAAATGCTGGGTGAG 324

```

```

RESULT 14
LOCUS AG186254
DEFINITION Pan troglodytes DNA, clone: RP43-060J11.T7, genomic survey
sequence.
ACCESSION AG186254
VERSION AG186254.1 GI:16715934
KEYWORDS GSS.
SOURCE Pan troglodytes (chimpanzee)
ORGANISM Pan troglodytes
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini;
Hominoidea; Pan.

```

```

REFERENCE 1
AUTHORS Fujiyama, A., Hattori, M., Toyoda, A., Taylor, T.D., Yada, T.,
Totoki, Y., Watanabe, H. and Sakaki, Y.
TITLE BAC end sequences of Library RPCI-43
JOURNAL Unpublished
REFERENCE 2 (bases 1 to 710)
AUTHORS Fujiyama, A., Hattori, M., Toyoda, A., Taylor, T.D., Yada, T.,
Totoki, Y., Watanabe, H. and Sakaki, Y.
TITLE Direct Submission
JOURNAL Submitted (02-AUG-2001) Asao Fujiyama, The Institute of Physical
and Chemical Research (RIKEN), Genomic Sciences Center (GSC);
1-7-22 Suehiro-chou, Tsukuba-shi, Ibaraki, Japan 305-0856, Japan
(E-mail: chimpes@gsr.riken.go.jp, URL: http://hgp.gsc.riken.go.jp/,
Tel: 81-45-503-9111, Fax: 81-45-503-9170)
COMMENT Clones are derived from the chimpanzee BAC library RPCI-43. This BAC
end was generated during the R&D process and may have higher chance
of clone tracking errors.
PRIMERS
Sequencing: T7
LIBRARY

```


Fri Feb 10 09:18:02 2006

Db 1452 GTGAAACCCCTCTCTAGTAAATAATAAAGTTAGCTGGACATGGTGGTAGGTGCCTGT 1393

866 AGTCCAGCTACTCAGGAGTCTGAGCAGGAGTATC 901

Qy 866 AGTCCCAGCTACTCAGAGTCTGAGCAGCAGGTATC 901
| | | | | | | | | | | | | | | | | | | | |
Db 1392 AATCCAGGCCACTCAGAGGCTGAGTAGGAGAATC 1357

Search completed: February 10, 2006, 00:58:12
Job time : 3374.06 secs

GenCore version 5.1.7
Copyright (c) 1993 - 2006 Bioacceleration Ltd.

OM nucleic - nucleic search, using sw model

Run on: February 9, 2006, 14:43:53 ; Search time 3492.02 Seconds
(without alignments)
11410.963 Million cell updates/sec

Title: US-10-607-806-1-G7328_COPY_7000_7700
Perfect score: 699.8
Sequence: 1 gtcgtgctactgctgtcag.....tcgcagcctcaacctctgag 701

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 5883141 seqs, 28421725653 residues
Total number of hits satisfying chosen parameters: 11766282

Minimum DB seq length: 0
Maximum DB seq length: 2000000000
Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : GenEmbl.*
1: gb.ba.*
2: gb.in.*
3: gb.env.*
4: gb.on.*
5: gb.ov.*
6: gb.pat.*
7: gb.ph.*
8: gb.pr.*
9: gb.ro.*
10: gb.sts.*
11: gb.sy.*
12: gb.un.*
13: gb.vi.*
14: gb.htg.*
15: gb.pl.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB	ID	Description
1	699.8	100.0	8368	8	AY438977	Homo sapiens phospholipase A2, group IB (pancreas) (PLA2G1B) gene, complete cds.
2	699.8	100.0	122302	8	AC003982	Homo sapiens
3	699.8	100.0	220384	14	AC078926	Homo sapiens
4	699.4	100.0	13612	6	AX377239	Sequence
5	380.8	54.4	3375	8	HUMPLA2A1	Sequence
6	355.8	50.8	357	10	G67590	Sequence
7	312	44.6	189729	14	AC073930	Sequence
8	211.4	30.2	227930	14	AC158058	Sequence
9	140.6	20.1	215223	14	AC127942	Sequence
10	140.6	20.1	238961	14	AC097575	Sequence
11	140.6	20.1	290053	14	AC096966	Sequence
12	130.8	18.7	418	6	CQ922146	Sequence
13	130.8	18.7	447	6	AX377240	Sequence
14	130.8	18.7	559	6	CQ723247	Sequence
15	130.8	18.7	562	6	AR380655	Sequence
16	130.8	18.7	562	6	AX332734	Sequence
17	130.8	18.7	562	6	AX333340	Sequence
18	130.8	18.7	562	6	AX333743	Sequence

19	130.8	18.7	562	8	HUMPLA2RA	M21054 Human lung
20	130.8	18.7	569	6	E02268	E02268 cDNA encodi
21	130.8	18.7	644	6	BD209899	BD209899 Human nuc
22	130.8	18.7	644	6	AX014213	AX014213 Sequence
23	130.8	18.7	654	6	CS071209	CS071209 Sequence
c 24	126.2	18.0	496	6	CS071001	CS071001 Sequence
25	126	18.0	10234	9	AF094611	AF094611 Mus muscu
c 26	126	18.0	183861	9	AC117735	AC117735 Mus muscu
27	126	18.0	215516	14	AC024694	AC024694 Mus muscu
28	126	18.0	238599	9	AC159539	AC159539 Mus muscu
29	96.8	13.8	562	4	DOGPHA2	M35301 Canine phos
30	96.8	13.8	567	4	DOGPPLA2	D00035 Canis sp. m
c 31	95.2	13.6	154362	14	AC069517	AC069517 Homo sapi
32	95.2	13.6	228098	8	AL596247	AL596247 Human DNA
33	92	13.1	542	9	RATPLA2	D00036 Rattus norv
34	90.8	13.0	546	4	RABPHOPO	L11899 Oryctolagus
35	90.4	12.9	527	9	AF187852	AF187852 Mus muscu
36	90.4	12.9	531	9	AF097637	AF097637 Mus muscu
37	90.4	12.9	542	9	AF094610	AF094610 Mus muscu
38	90.4	12.9	552	9	AF162712	AF162712 Mus muscu
39	89.8	12.8	369	6	E01687	E01687 cDNA encodi
40	89.8	12.8	546	4	BTPLA2R	Y00120 Bos taurus
41	89	12.7	173929	14	AC131928	AC131928 Homo sapi
c 42	89	12.7	189230	8	AC103974	AC103974 Homo sapi
c 43	89	12.7	198565	14	AC016750	AC016750 Homo sapi
44	88.2	12.6	483	4	SSPLA2R	Y00146 Porcine mRN
45	88.2	12.6	509	4	PIGPLA2RA	M21055 Pig pancrea

ALIGNMENTS

RESULT 1
AY438977
LOCUS
DEFINITION
ACCESSION
VERSION
KEYWORDS
SOURCE
ORGANISM
REFERENCE
AUTHORS
TITLE
JOURNAL
COMMENT
FEATURES

AY438977 Homo sapiens phospholipase A2, group IB (pancreas) (PLA2G1B) gene, complete cds.
AY438977
AY438977.1 GI:37953284
Homo sapiens (human)
Homo sapiens
Mammalia; Euthera; Euarchontoglires; Primates; Catarrhini; Homiidae; Homo.
1 (bases 1 to 8368)
Rieder,M.J., Livingston,R.J., Daniels,M.R., Chung,M.-W., Miyamoto,K.E., Nguyen,C.P., Nguyen,D.A., Poel,C.L., Robertson,P.D., Schackwitz,W.S., Sherwood,J.K., Witrak,L.A. and Nickerson,D.A.
Direct Submission
Submitted (16-OCT-2003) Genome Sciences, University of Washington, 1705 NE Pacific, Seattle, WA 98195, USA
To cite this work please use: NIEHS-SNPs, Environmental Genome Project, NIEHS ES15478, Department of Genome Sciences, Seattle, WA (URL: <http://egp.gs.washington.edu>).

source 1..8368
/organism="Homo sapiens"
/mol_type="genomic DNA"
/db_xref="taxon:9606"
repeat_region 101..257
/rpt_family="MIR"
/rpt_type="dispersed"
variation 205
/frequency="0.01"
/replace="a"
variation 306
/frequency="0.01"
/replace="a"
repeat_region 326..538
/rpt_family="Alu"
/rpt_type="dispersed"
misc_feature 354..496

Gene	/note="Region not scanned for variation"	variation	3630
	875. .6548		/gene="PLA2G1B"
mRNA	join(875. .944,2644. .2803,3603. .3730,6347. .6548)	variation	/frequency="0.10"
	/gene="PLA2G1B"		/replace="c"
	/product="phospholipase A2, group IB (pancreas)"		3702
CDS	join(911. .944,2644. .2803,3603. .3730,6347. .6471)	variation	/gene="PLA2G1B"
	/gene="PLA2G1B"		/frequency="0.15"
	/codon_start=1		/replace="a"
	/product="phospholipase A2, group IB (pancreas)"	variation	3768
	/protein_id="AA05441.1"		/gene="PLA2G1B"
	/db_xref="GI:3793285"	variation	/frequency="0.01"
	translation="MKLLVLAFLTVAAASGSPRAWQFRMKIKVIGSDPFLEY		/replace="t"
	NNYCYGLGGSGTPVDLDKCCQTHDNCYDQAKLDSCKFLDNPTHTYSYSCSGS	variation	3818
	AITCSKNKECEAFICNDRNAICFSKAPYKAHKNLDTKKYCSQ"		/gene="PLA2G1B"
repeat_region	1174. .1307	repeat_region	/frequency="0.01"
	/rpt_family="Alu"		/replace="a"
variation	1286	repeat_region	3846. .3964
	/gene="PLA2G1B"		/rpt_family="MIR"
	/frequency="0.01"		/rpt_type=dispersed
	/replace="g"		3977. .4280
repeat_region	1319. .1588	repeat_region	/rpt_family="Alu"
	/rpt_family="Alu"		/rpt_type=dispersed
variation	1343	repeat_region	4281. .4346
	/gene="PLA2G1B"		/rpt_family="L2"
	/frequency="0.03"		/rpt_type=dispersed
	/replace="a"		4347. .4634
variation	1593	variation	/rpt_family="Alu"
	/gene="PLA2G1B"		/rpt_type=dispersed
	/frequency="0.02"		4386
	/replace="t"		/gene="PLA2G1B"
repeat_region	1605. .1669	repeat_region	/frequency="0.08"
	/rpt_family="L2"		/replace="t"
	/rpt_type=dispersed		4635. .4723
misc_feature	1638. .2188	variation	/rpt_family="L2"
	/gene="PLA2G1B"		/rpt_type=dispersed
repeat_region	1670. .1975	repeat_region	5013
	/rpt_family="Alu"		/gene="PLA2G1B"
	/rpt_type=dispersed		/frequency="0.10"
repeat_region	1976. .2038	variation	/replace="a"
	/rpt_family="L2"		5103. .5404
	/rpt_type=dispersed		/rpt_family="Alu"
repeat_region	2086. .2165		/rpt_type=dispersed
	/rpt_family="L2"		5133
	/rpt_type=dispersed		/gene="PLA2G1B"
repeat_region	2194. .2319	variation	/frequency="0.01"
	/rpt_family="Alu"		/replace="c"
	/rpt_type=dispersed		5543
misc_feature	2319. .2453	variation	/gene="PLA2G1B"
	/gene="PLA2G1B"		/frequency="0.01"
variation	3027		/replace="c"
	/note="Region not scanned for variation"		5555
	/gene="PLA2G1B"		/gene="PLA2G1B"
	/frequency="0.25"		/frequency="0.01"
	/replace="t"		/replace="c"
repeat_region	3165. .3431	repeat_region	5556
	/rpt_family="Alu"		/gene="PLA2G1B"
	/rpt_type=dispersed		/frequency="0.15"
variation	3262	variation	/replace="g"
	/gene="PLA2G1B"		5577. .5629
	/frequency="0.01"		/rpt_family="L1"
	/replace="t"		/rpt_type=dispersed
variation	3321	variation	5580
	/gene="PLA2G1B"		/gene="PLA2G1B"
	/frequency="0.01"		/frequency="0.09"
	/replace="a"		/replace="g"
variation	3396	repeat_region	5675
	/gene="PLA2G1B"		/gene="PLA2G1B"
	/frequency="0.03"		/frequency="0.09"
	/replace="g"		/replace="g"
			5779. .6083
			/rpt_family="Alu"
			/rpt_type=dispersed
			5929


```
* 57227 80266: contig of 23040 bp in length
* 80267 80366: gap of unknown length
* 80367 109398: contig of 29032 bp in length
* 80368 109399: gap of unknown length
* 109399 149812: contig of 40314 bp in length
* 149812 149813: gap of unknown length
* 149813 220384: contig of 70472 bp in length.
FEATURES
    Location/Qualifiers
        1..220384
            /organism="Homo sapiens"
            /mol_type="genomic DNA"
            /db_xref="taxon:9606"
            /chromosome="12"
            /clone="RP11-836M11"
            2157..2256
                gap
            7113..7212
                gap
            /estimated_length=unknown
            13284..13383
                gap
            /estimated_length=unknown
            20661..20760
                gap
            /estimated_length=unknown
            34474..34573
                gap
            /estimated_length=unknown
            44395..44494
                gap
            /estimated_length=unknown
            57127..57226
                gap
            /estimated_length=unknown
            80267..80366
                gap
            /estimated_length=unknown
            109399..109498
                gap
            /estimated_length=unknown
            149813..149912
                gap
            /estimated_length=unknown
    ORIGIN
        Query Match      100.0%; Score 699.8; DB 14; Length 220384;
        Best Local Similarity 99.6%; Pred. No. 1.1e-179;
        Matches 698; Conservative 3; Mismatches 0; Indels 0; Gaps 0;

QY 1 GTCGTCTCACTGCTCCAGCTGGGTAAACAGAGCAACTCTGTCTCAAAAAAATAAATG 60
DB 71511 GTCGTCTCACTGCTCCAGCTGGGTAAACAGAGCAACTCTGTCTCAAAAAAATAAATG 71452

QY 61 CTTTCAATAAATATATATAAAGGACTTATATTTTCAAGCCATAGGATCATTTCTCC 120
DB 71451 CTTTCAATAAATATATATAAAGGACTTATATTTTCAAGCCATAGGATCATTTCTCC 71392

QY 121 TGAAGCATCTTGGCGAAGTCATCCCACTGTTCTTGAGAGTGGCGAGGTGAGGGCTGAC 180
DB 71391 TGAAGCATCTTGGCGAAGTCATCCCACTGTTCTTGAGAGTGGCGAGGTGAGGGCTGAC 71332

QY 181 CTATTGCTGCACTTACTCTATCTCAGCTGTCCTCCACATTTCCAGGTGCTGCCAGA 240
DB 71331 CTATTGCTGCACTTACTCTATCTCAGCTGTCCTCCACATTTCCAGGTGCTGCCAGA 71272

QY 241 CACATGCAACTGCTAYGACAGGCCAAGAGCTGGACAGCTGTAAATTTCTCTGGACA 300
DB 71271 CACATGCAACTGCTAYGACAGGCCAAGAGCTGGACAGCTGTAAATTTCTCTGGACA 71212

QY 301 MMCGGTACACCCACCATTTATCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 360
DB 71211 ACCCGTACACCCACCATTTATCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 71152

QY 361 GGTTTATCCCTCTCTGACCTATGAATTTCTAGTTGGTTCTCAGTAGGCCGGGGGGAATA 420
DB 71151 GGTTTATCCCTCTCTGACCTATGAATTTCTAGTTGGTTCTCAGTAGGCCGGGGGGAATA 71092

QY 421 ATAGTAAACAACAGCCATGATTAGTGTAAATTTTCTGGTCTGGGAGTGTCTCCCTTTA 480
DB 71091 ATAGTAAACAACAGCCATGATTAGTGTAAATTTTCTGGTCTGGGAGTGTCTCCCTTTA 71032

QY 481 ATCTCTAGAACCAACTATGGGATAGGTACAAATATCTCTACTTAACAGATAAGAAACT 540
DB 71032 ATCTCTAGAACCAACTATGGGATAGGTACAAATATCTCTACTTAACAGATAAGAAACT 540

Db 71031 ATCTCTAGAACCAACTATGGGATAGGTACAAATATCTCTACTTAACAGATAAGAAACT 70972
QY 541 GAGGCTCAGAAGCTGAGCTGATTTTCCCAAGATCACAACAGCTTGTAAAGTGGTACAGTTT 600
DB 70971 GAGGCTCAGAAGCTGAGCTGATTTTCCCAAGATCACAACAGCTTGTAAAGTGGTACAGTTT 70912
QY 601 GGGTTTTTTTTTTGTTGTTTGTAGACAGAGGCTTCTGCTCTGTCAACCCAGGCATGAGCAC 660
DB 70911 GGGTTTTTTTTTTGTTGTTTGTAGACAGAGGCTTCTGCTCTGTCAACCCAGGCATGAGCAC 70852
QY 661 AGTGGTCAACCATAGTCTCACTGCAGCTCAACCTCTCTGAG 701
DB 70851 AGTGGTCAACCATAGTCTCACTGCAGCTCAACCTCTCTGAG 70811

RESULT 4
AX377239 LOCUS AX377239 13612 bp DNA linear PAT 18-MAR-2002
DEFINITION Sequence 1 from Patent WO0212562.
ACCESSION AX377239
VERSION AX377239.1 GI:19573528
KEYWORDS
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini;
Homnidae; Homo.
REFERENCE
1 Kazemi, A., Kliem, S.B. and Koshy, B.
AUTHORS Haplotypes of the pla2g1b gene
TITLE Patent: WO 0212562-A 1 14-FEB-2002;
JOURNAL Genaissance Pharmaceuticals, Inc. (US)
FEATURES
    Location/Qualifiers
        1..13612
            /organism="Homo sapiens"
            /mol_type="unassigned DNA"
            /db_xref="taxon:9606"
            3845
                variation
            /note="PS1: polymorphic base G or A"
            3968
                variation
            /note="PS2: polymorphic base A or G"
            6060
                variation
            /note="PS2: polymorphic base A or G"
            6844
                variation
            /note="PS4: polymorphic base G or A"
            9531
                variation
            /note="PS5: polymorphic base G or A"
    ORIGIN
        Query Match      100.0%; Score 699.8; DB 6; Length 13612;
        Best Local Similarity 99.4%; Pred. No. 1.5e-179;
        Matches 697; Conservative 4; Mismatches 0; Indels 0; Gaps 0;

QY 1 GTCGTCTCACTGCTCCAGCTGGGTAAACAGAGCAACTCTGTCTCAAAAAAATAAATG 60
DB 6516 GTCGTCTCACTGCTCCAGCTGGGTAAACAGAGCAACTCTGTCTCAAAAAAATAAATG 6575

QY 61 CTTTCAATAAATATATATAAAGGACTTATATTTTCAAGCCATAGGATCATTTCTCC 120
DB 6576 CTTTCAATAAATATATATAAAGGACTTATATTTTCAAGCCATAGGATCATTTCTCC 6635

QY 121 TGAAGCATCTTGGCGAAGTCATCCCACTGTTCTTGAGAGTGGCGAGGTGAGGGCTGAC 180
DB 6636 TGAAGCATCTTGGCGAAGTCATCCCACTGTTCTTGAGAGTGGCGAGGTGAGGGCTGAC 6695

QY 181 CTATTGCTCTGCACTTACTCTCTATCTCAGCTGTCCTCCACATTTCCAGGTGCTGCCAGA 240
DB 6696 CTATTGCTCTGCACTTACTCTCTATCTCAGCTGTCCTCCACATTTCCAGGTGCTGCCAGA 6755

QY 241 CACATGCAACTGCTAYGACAGGCCAAGAGCTGGACAGCTGTAAATTTCTGCTGGACA 300
DB 6756 CACATGCAACTGCTAYGACAGGCCAAGAGCTGGACAGCTGTAAATTTCTGCTGGACA 6815
```

```
QY 301 MMCCGTACACCCACACCTATTATCTGCTCTGGCTCGGCAATCATCCTGTAGCAGTA 360
Db 6816 ACCCGTACACCCACACCTATTATCTGCTCTGGCTCGGCAATCATCCTGTAGCAGTA 6875
QY 361 GGTATTATCCCTTCCCTTGACCTATGAAATCTAGTTGGTTCTCAGTAGGCGGGGGGAATA 420
Db 6876 GGTATTATCCCTTCCCTTGACCTATGAAATCTAGTTGGTTCTCAGTAGGCGGGGGGAATA 6935
QY 421 ATAGTAACAACAGCCATGATTTAGTGTAAATTTTCTGGTCTCGGCAAGTGTCTCTTTA 480
Db 6936 ATAGTAACAACAGCCATGATTTAGTGTAAATTTTCTGGTCTCGGCAAGTGTCTCTTTA 6995
QY 481 ATCTCAGAACACACTATGGATAGGTACAAATTAATCTCTAAACAGATAAAGAACT 540
Db 6996 ATCTCAGAACACACTATGGATAGGTACAAATTAATCTCTAAACAGATAAAGAACT 7055
QY 541 GAGGCTCAGAAGGCTGAGCTATTGCCCCAAGATCACAGCTTCTTAAGTGTGACAGTTT 600
Db 7056 GAGGCTCAGAAGGCTGAGCTATTGCCCCAAGATCACAGCTTCTTAAGTGTGACAGTTT 7115
QY 601 GGGTTTTTTTTTGTGTTTGTAGAGACAGGGTCTTGCTCTGTCACCCAGGCATGAGCAC 660
Db 7116 GGGTTTTTTTTTGTGTTTGTAGAGACAGGGTCTTGCTCTGTCACCCAGGCATGAGCAC 7175
QY 661 AGTGGTGAACCATAGGTCACTGCAGCCTCAACCTCCCTGAG 701
Db 7176 AGTGGTGAACCATAGGTCACTGCAGCCTCAACCTCCCTGAG 7216

RESULT 5
HUMPLA2A1
LOCUS Human pancreatic phospholipase A-2 (PLA-2) gene, exons 1 to 3.
DEFINITION Human pancreatic phospholipase A-2 (PLA-2) gene, exons 1 to 3.
ACCESSION M22970 M14965
VERSION M22970.1 GI:190008
KEYWORDS phospholipase A2.
SEGMENT 1 of 2
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini;
Hominidae; Homo.
REFERENCE 1 (bases 1 to 3375)
AUTHORS Seilhamer,J.J., Randall,T.L., Yamanaka,M. and Johnson,L.K.
TITLE Pancreatic phospholipase A2: isolation of the human gene and cDNAs
JOURNAL from porcine pancreas and human lung
PUBMED DNA 5 (6), 519-527 (1986)
COMMENT 3028739
FEATURES
source
    source text: Homo sapiens pulverized liver DNA.
    Location/Qualifiers
        1..3375
            /organism="Homo sapiens"
            /mol_type="genomic DNA"
            /db_xref="taxon:9606"
            /map="12q23-qter"
            /tissue type="pulverized liver"
            <537..570
            /gene="PLA2A"
            /note="G00-120-715"
            /number=1
            571..2262
            /gene="PLA2A"
            /note="G00-120-715"
            /number=1
            2263..2422
            /gene="PLA2A"
            /note="G00-120-715"
            /number=2
            2423..3222
            /gene="PLA2A"
            /note="G00-120-715"
            /number=2

exon
    3223..3350
        /gene="PLA2A"
        /note="G00-120-715"
        /number=3
        3351..3375
        /gene="PLA2A"
        /note="G00-120-715"
        /number=3
    5 bp upstream of PstI site.

intron
    Query Match 54.4%; Score 380.8; DB 8; Length 3375;
    Best Local Similarity 99.2%; Pred. No. 1e-92;
    Matches 379; Conservative 3; Mismatches 0; Indels 0; Gaps 0;

QY 1 GTCTGTCACTGTCTCCAGCCTGGGTAAACAGAGCAACTCTGTCTCAAAAAAATAATG 60
Db 2994 GTCTGTCACTGTCTCCAGCCTGGGTAAACAGAGCAACTCTGTCTCAAAAAAATAATG 3053
QY 61 CTTTCAATAAATATATGATAAAAGACTTATATTTTCAAGCCATAGGATCATTTCTCC 120
Db 3054 CTTTCAATAAATATATGATAAAAGACTTATATTTTCAAGCCATAGGATCATTTCTCC 3113
QY 121 TGAAGCATCTTGGCGAAGTATCCCACTGTTCTCGAGAGTGGGAGGTGAGGGCTGAC 180
Db 3114 TGAAGCATCTTGGCGAAGTATCCCACTGTTCTCGAGAGTGGGAGGTGAGGGCTGAC 3173
QY 181 CTATTGCTCTGCACTTACTCTCTATCTCAGCTGCCCTCCACTTTCCAGGTGCTGCAGA 240
Db 3174 CTATTGCTCTGCACTTACTCTCTATCTCAGCTGCCCTCCACTTTCCAGGTGCTGCAGA 3233
QY 241 CACATGCAACTGCTATGAYACCAGGCCAAGAGTGGACAGCTGATAAATTTCTCTTGACA 300
Db 3234 CACATGCAACTGCTATGAYACCAGGCCAAGAGTGGACAGCTGATAAATTTCTCTTGACA 3293
QY 301 MMCCGTACACCCACACCTATTATCTGCTCTGGCTCGGCAATCATCCTGTAGCAGTA 360
Db 3294 ACCCGTACACCCACACCTATTATCTGCTCTGGCTCGGCAATCATCCTGTAGCAGTA 3353
QY 361 GGTATTATCCCTTCCCTTGACCTA 382
Db 3354 GGTATTATCCCTTCCCTTGACCTA 3375

RESULT 6
G67590
LOCUS csnpla2glb-pcr3-1 Human Homo sapiens STS genomic, sequence tagged
DEFINITION site.
ACCESSION G67590
VERSION G67590.1 GI:12642952
KEYWORDS STS.
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini;
Hominidae; Homo.
REFERENCE 1 (bases 1 to 357)
AUTHORS Wong,G.K.S., Yu,J., Yang,Z., Passey,D., Kibukawa,M., Paddock,M. and
Olson,M.
TITLE Gene based polymorphism discovery
JOURNAL Unpublished (2000)
COMMENT
    Contact: Gane Ka-Shu Wong
    Genome Center
    University of Washington
    Fluke Hall, Box 352145, Seattle, WA 98195, USA
    Tel: 206/685-7348
    Fax: 206/685-7344
    Email: gkw@u.washington.edu
    Primer A: AAGCATCTTGGCGAAGTCAT
    Primer B: AAAGGAGACACTGCCAGAA
    STS size: 357
    PCR Profile:
```

Presoak:92 degrees C for 1.00 minute

Denaturation:92 degrees C for 0.16 minute

Annealing:60-65 degrees C for 0.50 minute

Polymerization:72 degrees C for 1.00 minute

PCR Cycles:35

Thermal Cycler:Perkin Elmer TC

Protocol:

Template: 7 ng

Primer: each 0.5 uM

dNTPs: each 100 uM

Taq Polymerase: 0.025 units/ul

Total Vol: 10 ul

Buffer:

MgCl2: 1.5 mM

KCl: 50 mM

Tris-HCl: 10 mM

pH: 8.3

Location/Qualifiers

1. .357

/organism="Homo sapiens"

/mol_type="genomic DNA"

/db_xref="taxon:9606"

/map="12q23-q24.1"

/clone_lib="Human"

1. .357

/gene="PLA2G1B"

1. .357

/gene="PLA2G1B"

1. .20

/gene="PLA2G1B"

complement (338.. .357)

gene

STS

primer_bind

primer_bind

ORIGIN

Query Match

Best Local Similarity 50.8%; Score 355.8; DB 10; Length 357;

Matches 354; Conservative 3; Mismatches 0; Indels 0; Gaps 0;

QY

123

AGCATCTTGGGAAGTCATCCCACTGTTCTCTGAGAGTGGCAGGTGAGGGCTGACCT

182

Db

1

AAGCATCTTGGGAAGTCATCCCACTGTTCTCTGAGAGTGGCAGGTGAGGGCTGACCT

60

QY

183

ATTGCTCTGCACCTACTCTCTATCTCAGCTGTCCCTCCCACTTCCAGGTGCTGCCAGACA

242

Db

61

ATTGCTCTGCACCTACTCTCTATCTCAGCTGTCCCTCCCACTTCCAGGTGCTGCCAGACA

120

QY

243

CATGACAACTGTATGACACCGCCAGAGCTGGACAGCTGTAAATTTCTGCTGGACAMM

302

Db

121

CATGACAACTGTATGACACCGCCAGAGCTGGACAGCTGTAAATTTCTGCTGGACAA

180

QY

303

CCGTACACCCACCACTTATCATCTGCTGCTGGCTCGGCAATCACCTGTAGCAGTAGG

362

Db

181

CCGTACACCCACCACTTATCATCTGCTGCTGGCTCGGCAATCACCTGTAGCAGTAGG

240

QY

363

TTTATCCCTTCTTGGACCTATGAATTTCTAGTTGGTTCTCAGTAGGCCGGGGGAAATAAT

422

Db

241

TTTATCCCTTCTTGGACCTATGAATTTCTAGTTGGTTCTCAGTAGGCCGGGGGAAATAAT

300

QY

423

AGTAAACAACAGCCATGATTTAGTTGTTAAATTTCTGGTCTGGGAGTGTCTCCCTTT

479

Db

301

AGTAAACAACAGCCATGATTTAGTTGTTAAATTTCTGGTCTGGGAGTGTCTCCCTTT

357

RESULT 7

AC073930

LOCUS

DEFINITION

AC073930

ACCESSION

AC073930

VERSION

HTG; HTGS_PHASE1; HTGS_DRAFT; HTGS_FULLTOP.

KEYWORDS

HTG; HTGS_PHASE1; HTGS_DRAFT; HTGS_FULLTOP.

SOURCE

Homo sapiens

ORGANISM

Homo sapiens

REFERENCE

AUTHORS

1 (bases 1 to 189729)

Muzny,D.M., Adams,C., Adio-Oduola,B., Ali-osman,F.R., Allen,C., Alsbrooks,S.L., Amaratunge,H.C., Are,J.R., Ayele,M., Banks,T., Barbaria,J., Benton,J., Binnig,K., Blankenburg,K., Bonnin,D., Bouck,J., Bowie,S., Brieva,M., Brown,E., Brown,M., Bryant,N.P., Buhay,C., Burch,P., Burkett,C., Burrell,K.L., Byrd,N.C., Carron,T.F., Carter,M., Cavazos,S.R., Chacko,J., Chavez,D., Chen,G., Chen,R., Chen,Z., Chowdhry,I., Christopoulos,C., Cleveland,C.D., Cox,C., Coyne,M.D., Dathorne,S.R., David,R., Davila,M.L., Davis,C., Davy-Carroll,L., Dederich,D.A., Delaney,K.R., Delgado,O., Denn,A.L., Ding,Y., Dinh,H.H., Earthwaite,K.J., Draper,H., Dugan-Rocha,S., Durbin,K.J., Earhart,C., Edgar,D., Edwards,C.C., Elhaj,C., Escotto,M., Falls,T., Ferraguto,D., Flagg,N., Ford,J., Foster,P., Frantz,P., Gabisi,A., Gao,J., Garcia,A., Garner,T., Garza,N., Gill,R., Gorrell,J.H., Guevara,W., Gunaratne,P., Hale,S., Hamilton,K., Harris,C., Harris,K., Hart,M., Havlak,P., Hawes,A., Hernandez,J., Hernandez,O., Hodgson,A., Hogues,M., Holloway,C., Hollins,B., Homs,F., Howard,S., Huber,J., Hulyk,S., Hume,J., Jackson,L.E., Jacobson,B., Jia,Y., Johnson,R., Jolivet,S., Joudah,S., Karlsson,E., Kelly,S., Khan,U., King,L., Korvah,J., Kovar,C., Kratovic,J., Kureshi,A., Landry,N., Leal,B., Lewis,L.C., Lewis,S., Li,J., Li,Z., Lichtarge,O., Lieu,C., Liu,J., Liu,W., Loulseghe,H., Lozado,R.J., Lu,X., Lucier,A., Lucier,R., Luna,R., Ma,J., Maheshwari,M., Mapua,P., Martin,R., Martindale,A., Martinez,E., Massey,E., Mawhiney,E., McLeod,M.P., Meador,M., Mei,G., Metzker,M., Miner,G., Miner,Z., Mitchell,T., Mohabbat,K., Morgan,M., Morris,S., Moser,M., Neal,D., Newton,J., Newton,N., Nguyen,A., Nguyen,N., Nguyen,N., Nickerson,E., Nwokwenkwo,S., Oguh,M., Okwuonu,G., Oragunye,N., Owiedo,R., Pace,A., Payton,B., Peery,J., Perez,L., Peters,L., Pickens,R., Primus,E., Fu,L.L., Quiles,M., Ran,Y., Rives,M., Rojas,A., Rojubokan,I., Rolfe,M., Ruiz,S., Saverly,G., Scherer,S., Scott,G., Shen,H., Shoohtari,N., Sisson,I., Sodergren,E., Sonalike,T., Sparks,A., Stanley,H., Stone,H., Sutton,A., Svatek,A., Taber,P., Tamerisa,A., Tamerisa,K., Tang,H., Tansey,J., Taylor,C., Taylor,T., Telford,B., Thomas,N., Thomas,S., Uman,I., K., Vasquez,L., Vera,V., Villalon,D., Vinson,R., Wang,Q., Wang,S., Ward-Moore,S., Warren,R., Washington,C., Watlington,S., Williams,G., Williamson,A., Wleczyk,R., Wooden,S., Worley,K., Wu,C., Wu,Y., Wu,Y.F., Zhou,J., Zorrilla,S., Nelson,D., Weinstock,G. and Gibbs,R.

Unpublished

2 (bases 1 to 189729)

Worley,K.C.

Direct Submission

Submitted (05-JUL-2000) Human Genome Sequencing Center, Department of Molecular and Human Genetics, Baylor College of Medicine, One Baylor Plaza, Houston, TX 77030, USA

3 (bases 1 to 189729)

Worley,K.C.

Direct Submission

Submitted (01-MAR-2002) Human Genome Sequencing Center, Department of Molecular and Human Genetics, Baylor College of Medicine, One Baylor Plaza, Houston, TX 77030, USA

On Feb 27, 2002 this sequence version replaced gi:14861669.

----- Genome Center

Center: Baylor College of Medicine

Center code: BCM

Web site: <http://www.hgsc.bcm.tmc.edu/>

Contact: hgsc-help@bcm.tmc.edu

----- Project Information

Center project name: HBWS

Center clone name: RP11-144B2

----- Summary Statistics

Sequencing vector: Plasmid; M77789

Chemistry: Dye-terminator Big Dye; 100% of reads

Assembly program: Phrap; version 0.990329

Consensus quality: 191640 bases at least Q40

Consensus quality: 195135 bases at least Q30
Consensus quality: 197786 bases at least Q20
Estimated insert size: 193340; sum-of-contigs estimation
Quality coverage: 0x in Q20 bases; agarose-fp estimation
Quality coverage: 11.7x in Q20 bases; sum-of-contigs estimation

* NOTE: Estimated insert size may differ from sequence length
(see http://www.hgsc.bcm.tmc.edu/docs/Genbank_draft_data.html).
* NOTE: This is a 'working draft' sequence. It currently
* consists of 8 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.
* 1 64773: contig of 64773 bp in length
* 64774 64873: gap of unknown length
* 64874 99097: contig of 34224 bp in length
* 99098 99197: gap of unknown length
* 99198 135331: contig of 36134 bp in length
* 135332 135431: gap of unknown length
* 135432 161344: contig of 25913 bp in length
* 161345 161444: gap of unknown length
* 161445 171231: contig of 9787 bp in length
* 171232 171331: gap of unknown length
* 171332 180922: contig of 9591 bp in length
* 180923 181022: gap of unknown length
* 181023 187271: contig of 6249 bp in length
* 187272 187371: gap of unknown length
* 187372 189729: contig of 2358 bp in length.
* 189729 Location/Qualifiers

FEATURES
source

1. 189729
/organism="Homo sapiens"
/mol_type="genomic DNA"
/db_xref="taxon:9606"
/chromosome="12q"
/clone="RP11-144B2"
64774. 64873
/estimated_length=unknown
99098. 99197
/estimated_length=unknown
135332. 135431
/estimated_length=unknown
161345. 161444
/estimated_length=unknown
171232. 171331
/estimated_length=unknown
180923. 181022
/estimated_length=unknown
187272. 187371
/estimated_length=unknown

ORIGIN

Query Match 44.6%; Score 312; DB 14; Length 189729;
Best Local Similarity 100.0%; Pred. No. 5.4e-74;
Matches 312; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 390 TAGTTGGTCTCAGTAGCCCGGGGAAATATAGTACACACCCATGATTAGTGTGA 449
DB 171332 TAGTTGGTCTCAGTAGCCCGGGGAAATATAGTACACACCCATGATTAGTGTGA 171391
QY 450 ATTTCTTGGTCTTGGGCGAGTGCTCTTTTAACTCTCAGAACACACTATGGGATAGTA 509
DB 171392 ATTTCTTGGTCTTGGGCGAGTGCTCTTTTAACTCTCAGAACACACTATGGGATAGTA 171451
QY 510 CAATTATCTCTTAACAGATAAGAAAACCTGAGGCTCAGAAGCTGAGCTATTGGCCCA 569
DB 171452 CAATTATCTCTTAACAGATAAGAAAACCTGAGGCTCAGAAGCTGAGCTATTGGCCCA 171511
QY 570 AGATCACACAGCTTGTAAAGTGGTGACAGTTGGGTTTTTTTTTTGTTGTTGTTAGAGACA 629
DB 171512 AGATCACACAGCTTGTAAAGTGGTGACAGTTGGGTTTTTTTTTTGTTGTTGTTAGAGACA 171571

QY 630 GGGTCTTGCTCTGTGTCACCCAGGATGACACAGTGGTGCAACCATAGTCTCACTGAGCT 689
DB 171572 GGGTCTTGCTCTGTGTCACCCAGGATGACACAGTGGTGCAACCATAGTCTCACTGAGCT 171631
QY 690 CAACCTCTCTGAG 701
DB 171632 CAACCTCTCTGAG 171643

RESULT 8
AC158058/c

LOCUS AC158058 227930 bp DNA linear HTG 01-JUL-2005
DEFINITION Bos taurus clone CH240-58P17, WORKING DRAFT SEQUENCE, 27 unordered pieces.
ACCESSION AC158058
KEYWORDS HTG; HTGS_PHASE1; HTGS_DRAFT; HTGS_FULLTOP.
SOURCE Bos taurus (cow)
ORGANISM Bos taurus

ORGANISM

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Laurasiatheria; Cetartiodactyla; Ruminantia; Pecora; Bovidae; Bovinae; Bos.
1 (bases 1 to 227930)

REFERENCE
AUTHORS

Muzny, D. Marie., Metzker, M. Lee., Abramson, S., Adams, C., Alder, J., Allen, C., Allen, H., Alsbrooks, S., Amin, A., Anguiano, D., Anyalebechi, V., Aoyagi, A., Ayodeji, M., Baca, E., Baden, H., Baldwin, D., Bandaranaike, D., Barber, M., Barnstead, M., Benahmed, F., Biewald, K., Blair, J., Blankenburg, K., Blyth, P., Brown, M., Bryant, N., Buhay, C., Burch, P., Burrell, K., Calderon, E., Cardenas, V., Carter, K., Cavazos, I., Ceasar, H., Chen, A., Chacko, J., Chavez, D., Chen, G., Chen, R., Chen, Y., Chen, Z., Cleaveland, C., Cockrell, R., Cox, C., Coyle, M., Cree, A., D'Souza, L., Davila, M. L., Davis, C., Davy-Carroll, L., De Anda, C., Dederich, D., Delgado, O., Denson, S., Deramo, C., Ding, Y., Dinh, H., Divya, K., Draper, H., Dugan-Rocha, S., Dunn, A., Durbin, K., Duval, B., Eaves, K., Egan, A., Escotto, M., Eugene, C., Evans, C. A., Falls, T., Fan, G., Fernandez, S., Finley, M., Flagg, N., Forbes, L., Foster, M., Foster, P., Fraser, C. M., Gabisi, A., Ganta, R., Garcia, A., Garner, T., Garza, M., Gebregorgis, E., Geer, K., Gill, R., Grady, M., Guerra, W., Guevara, W., Gunaratne, P., Haaland, W., Hamill, C., Hamilton, C., Hamilton, K., Harvey, Y., Havlak, P., Hawes, A., Henderson, N., Hernandez, J., Hollins, B., Hines, S., Hladun, S. L., Hodgson, A., Hoques, M., Jackson, L., Jacob, L., Jiang, H., Johnson, B., Johnson, R., Jolivet, A., Karpathy, S., Kelly, S., Kelly, S., Khan, Z., King, L., Kovar, C., Kowis, C., Kraft, C. L., Lebow, H., Levan, J., Lewis, L., Li, Z., Liu, J., Liu, J., Liu, W., Liu, Y., London, P., Longacre, S., Lopez, J., Lorensuhewa, L., Loulseghe, H., Lozado, R. J., Lu, X., Ma, J., Maheshwari, M., Mahindartne, M., Mahmoud, M., Malloy, K., Mangum, A., Mangum, B., Mapua, P., Martin, K., Martin, R., Martinez, E., Mawhiney, S., McLeod, M. P., McNeill, T. Z., Meenen, E., Milosavljevic, A., Miner, G., Minja, E., Montemayor, J., Moore, S., Morgan, M., Morris, K., Morris, S., Munidasa, M., Murphy, M., Nair, L., Narkervis, C., Neal, D., Newton, N., Nguyen, N., Norris, S., Nwaokelemeh, O., Okwuonog, G., Olarnpungsoo, A., Pal, S., Parks, K., Pasternak, S., Paul, H., Perez, A., Perez, L., Pfannkuch, C., Plopper, F., Poindexter, A., Popovic, D., Primus, E., Pu, L., L., Puzo, M., Quiroz, J., Rachlin, E., Reeves, K., Regier, M. A., Reigh, R., Reilly, B., Reilly, M., Ren, Y., Reuter, M., Richards, S., Riggs, F., Rives, C., Rodkey, T., Rojas, A., Rose, M., Rose, R., Ruiz, S. J., Sanders, W., Savary, G., Scherer, S., Scott, G., Shatsman, S., Shen, H., Shetty, J., Shvartsbeyn, A., Sison, I., Sitter, C. D., Smajis, D., Sneed, A., Sodergren, E., Song, X., Z., Sorelle, R., Sosa, J., Steinkle, M., Strong, R., Sutton, A., Svatek, A., Tabor, P., Taylor, C., Taylor, T., Thomas, N., Thomas, S., Tingey, A., Trejos, Z., Usmani, K., Valas, R., Vera, V., Villaseana, D., Waldron, L., Walker, B., Wang, J., Wang, Q., Wang, S., Warren, J., Warren, R., Wei, X., White, F., Williams, G., Willson, R., Wleczek, R., Wooden, H., Worley, K., Wright, D., Wright, R., Wu, J., Yakub, S., Yen, J., Yoon, L., Yoon, V., Yu, P., Zhang, J., Zhou, J., Zhou, X., Zhao, S., Dunn, P., von Niederhausern, A., Weiss, R., Smith, D. R., Holt, R. A., Smith, H. O., Weinstock, G. and Gibbs, R. A.

TITLE
JOURNAL
REFERENCE
AUTHORS
TITLE
JOURNAL

REFERENCE
AUTHORS
TITLE
JOURNAL

COMMENT

Direct Submission
 Unpublished
 2 (bases 1 to 227930)
 Worley, K.C.
 Direct Submission
 Submitted (04-MAR-2005) Human Genome Sequencing Center, Department
 of Molecular and Human Genetics, Baylor College of Medicine, One
 Baylor Plaza, Houston, TX 77030, USA
 3 (bases 1 to 227930)
 Cow Genome Sequencing Consortium.
 Direct Submission
 Submitted (01-JUL-2005) Human Genome Sequencing Center, Department
 of Molecular and Human Genetics, Baylor College of Medicine, One
 Baylor Plaza, Houston, TX 77030, USA
 The sequence in this assembly is a combination of BAC based reads
 and whole genome shotgun sequencing reads assembled using Atlas
 (http://www.hgsc.bcm.tmc.edu/projects/rat/). Each contig described
 in the feature table below represents a scaffold in the Atlas
 assembly (a 'contig-scaffold'). Within each contig-scaffold,
 individual sequence contigs are ordered and oriented, and separated
 by sized gaps filled with Ns to the estimated size. The sequence
 may extend beyond the ends of the clone and there may be sequence
 contigs within a contig-scaffold that consist entirely of whole
 genome shotgun sequence reads. Both end sequences and whole genome
 shotgun sequence only contigs will be indicated in the feature
 table.

----- Genome Center
 Center: Baylor College of Medicine
 Center code: BCM
 Web site: http://www.hgsc.bcm.tmc.edu/
 Contact: hgsc-help@bcm.tmc.edu
 ----- Project Information
 Center project name: PDGU
 Center clone name: CH240-58P17
 ----- Summary Statistics
 Assembly program: Atlas 3.0;
 Consensus quality: 218051 bases at least Q40
 Consensus quality: 219911 bases at least Q30
 Consensus quality: 221398 bases at least Q20
 Estimated insert size: 220393; sum-of-contigs estimation
 Quality coverage: 7x in Q20 bases; sum-of-contigs estimation

* NOTE: Estimated insert size may differ from sequence length
 (see http://www.hgsc.bcm.tmc.edu/docs/Genbank_draft_data.html).
 * NOTE: This is a 'working draft' sequence. It currently
 * consists of 27 contigs. The true order of the pieces
 * is not known and their order in this sequence record is
 * arbitrary. Gaps between the contigs are represented as
 * runs of N, but the exact sizes of the gaps are unknown.
 * This record will be updated with the finished sequence
 * as soon as it is available and the accession number will
 * be preserved.

1	6423:	contig of 6423 bp in length
*	6424	6574: gap of 151 bp
*	6575	11190: contig of 4616 bp in length
*	11191	11364: gap of 174 bp
*	11365	32898: contig of 21534 bp in length
*	32899	32899: gap of unknown length
*	32999	57556: contig of 24558 bp in length
*	57557	57606: gap of 50 bp
*	57607	70177: contig of 12571 bp in length
*	70178	70227: gap of 50 bp
*	70228	80485: contig of 10258 bp in length
*	80486	80535: gap of 50 bp
*	80536	86613: contig of 6078 bp in length
*	86614	86663: gap of 50 bp
*	86664	105235: contig of 18572 bp in length
*	105236	105298: gap of 63 bp
*	105299	114241: contig of 8943 bp in length
*	114242	114291: gap of 50 bp
*	114292	115475: contig of 1184 bp in length
*	115476	115525: gap of 50 bp

Db 32892 TGTCTTCAAGCCACAGGACCACTACCTCTCAGAGCAGGTGATCTTAAGACAGCCACCTG 32893
QY 152 TTCCTGAGAGTGGGCGAGGTGAGGCTGACCTATTGCTCTGCACTTACTCTTCTATCTCAGCT 211
Db 32832 CTCC--AAGGAGCAAGTTCAGGCTGGCGTCTGCTCCCTACCCATGCTCTCTGCTCAGCT 32776
QY 212 --GTCCCTCCCACTTTCAGGCTGCTGCGAGCACAATGACCACTGCTGAYGACAGGCCAAG 269
Db 32775 CCGTTCCTCCCTCTCAGGCTGCTGCCAGAGCGACCACTGCTACAAACAGGCCAAG 32716
QY 270 AAGCTGGACAGCTGTAATTTCTGCTGGACAMCCGTACCCACCACTTATCTATCTGCG 329
Db 32715 AAAGCTGGACAGCTGTAATTTCTGCTGGACAMCCGTACCCACCACTTATCTATCTGCG 32656
QY 330 TGCTCTGCTCGGCAATCACTGTAGCTAGTGGTTTATCTCTCTTCCCTGACCTATGAATTC 389
Db 32655 TGTCTTAACAATGAGATCACCTGACGAGTAGTGGTTTATCTCTCTTGGT----- 32608
QY 390 TAGTGGTCTCAGTAGCGCGGGGGAATATAGTATACACAGCCATGATTTAGTGTGA 449
Db 32607 --GCTCTTGTAGCGGGCATGGAGGGAATATAGTATGATAGCCATCACTGACTGATC 32550
QY 450 ATTTCTTGTGTTCTGGGAGTCTCTCTTAACTCTCAGACACACTATGGGTAGGTA 509
Db 32549 ATTTCTTGTGTTCTGGGAGTCTCTCTTAACTCTCAGACACACTATGGGTAGGTA 32490
QY 510 C---AATATCTCTCTTAACTCTCAGACACACTATGGGTAGGTTTATGGC 566
Db 32489 CTGTTATTTATCCCATTTTACAGATGAGGCACTGAGGCTCAGAGGCTGACAAATGT 32430
QY 567 CCAAGATCACAGCTTGTAGTGGTGCAGTTCAGTTGGGTTTT 607
Db 32429 CCAAGGGCACACAGTCCCATTAATGTTGGTGTGTTGTTTT 32389
RESULT 9
AC127942/c
LOCUS
DEFINITION Rattus norvegicus clone CH230-142L1, WORKING DRAFT SEQUENCE, 4
ACCESSION AC127942
VERSION AC127942.2 GI:23907779
KEYWORDS HTG; HTGS_PHASE1; HTGS_DRAFT; HTGS_FULLTOP.
SOURCE Rattus norvegicus
ORGANISM Rattus norvegicus
REFERENCE
AUTHORS Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Euarchontoglires; Glires; Rodentia;
Sciurognathi; Muroidae; Muridae; Murinae; Rattus.
Muzny,D,Marie., Metzker,M, Lee., Abramzon,S., Adams,C., Alder,J.,
Allen,C., Allen,H., Albrechts,S., Amin,A., Anguiano,D.,
Anyalebechi,V., Aoyagi,A., Ayodeji,M., Baca,E., Baden,H.,
Baldwin,D., Bandaranaike,D., Barber,M., Barnstead,M., Benahmed,F.,
Blawie,K., Blair,J., Blankenburg,K., Blyth,P., Brown,M.,
Bryant,N., Buhay,C., Burch,P., Burrell,K., Calderon,E.,
Cardenas,V., Carter,K., Cavazos,I., Cesar,H., Center,A.,
Chacko,J., Chavez,D., Chen,G., Chen,R., Chen,Y., Chen,Z., Chu,J.,
Cleveland,C., Cockrell,R., Cox,C., Coyle,M., Cree,A., D'Souza,L.,
Davila,M.L., Davis,C., Davy-Carroll,L., De Anda,C., Dederich,D.,
Delgado,O., Denson,S., Deramo,C., Ding,Y., Dinh,H., Divya,K.,
Draper,H., Dugan-Rocha,S., Dunn,A., Durbin,K., Duval,B., Evans,K.,
Egan,A., Escotto,M., Eugene,C., Evans,C.A., Falls,T., Fan,G.,
Fernandez,S., Finley,M., Flagg,N., Forbes,L., Foster,M., Foster,P.,
Fraser,C.M., Gabisi,A., Ganta,R., Garcia,A., Garner,T., Garza,M.,
Gebregorgis,E., Geer,K., Gill,R., Grady,M., Guerra,W., Guevara,W.,
Gunaratne,P., Haaland,W., Hamill,C., Hamilton,C., Hamilton,K.,
Harvey,Y., Havlak,P., Haves,A., Henderson,N., Hernandez,J.,
Hernandez,R., Hines,S., Hladun,S.L., Hodgson,A., Hoques,M.,
Hollins,B., Howells,S., Hulyk,S., Hume,J., Idlebird,D., Jackson,A.,
Jackson,L., Jacob,L., Jiang,H., Johnson,B., Johnson,R., Jolivet,A.,
Karpachy,S., Kelly,S., Kelly,S., Khan,Z., King,L., Kovar,C.,
Kowis,C., Kraft,C.L., Lebow,H., Levan,J., Lewis,L., Li,Z., Liu,J.,
Liu,J., Liu,W., Liu,Y., London,P., Longacre,S., Lopez,J.,

Lorensuhewa,L., Loulseged,H., Lozado,R.J., Lu,X., Ma,J.,
Maheshwari,M., Mahindartne,M., Mahmoud,M., Malloy,K., Mangum,A.,
Mangum,B., Mapua,P., Martin,K., Martin,R., Martinez,E.,
Mawhiney,S., McLeod,M.P., McNeill,T.Z., Meenen,E.,
Miloavljevic,A., Miner,G., Minja,S., Montemayor,J., Moore,S.,
Morgan,M., Morris,K., Morris,S., Munidasa,M., Murphy,M., Nair,L.,
Nankervis,C., Neal,D., Newton,N., Nguyen,N., Norris,S.,
Nwaokemele,O., Okwuonu,G., Olarnpunsagoon,A., Pal,S., Parks,K.,
Pasternak,S., Paul,H., Perez,A., Perez,L., Pfannkoch,C.,
Plopper,F., Poindexter,A., Popovic,D., Primus,E., Pu,L.,
Puzo,M., Quiroz,J., Rachin,E., Reeves,K., Regier,M.A., Reigh,R.,
Reilly,B., Reilly,M., Ren,Y., Reuter,M., Richards,S., Riggs,F.,
Rives,C., Rodkey,T., Rojas,A., Rose,M., Rose,R., Ruiz,S.J.,
Sanders,W., Savery,G., Scherer,S., Scott,G., Shatsman,S., Shen,H.,
Shetty,J., Shvartsbeyn,A., Sisson,I., Sitter,C.D., Smajic,D.,
Sneed,A., Sodergren,E., Song,X.-Z., Sorelle,R., Sosa,J.,
Steinle,M., Strong,R., Sutton,A., Svatek,A., Taber,P., Taylor,C.,
Taylor,T., Thomas,N., Thomas,S., Tingey,A., Trejos,Z., Usmani,K.,
Valas,R., Vera,V., Villasana,D., Waldron,L., Walker,B., Wang,J.,
Wang,Q., Wang,S., Warren,J., Warren,R., Wei,X., White,F.,
Williams,G., Willson,R., Wleczek,R., Wooden,H., Worley,K.,
Wright,D., Wright,R., Wu,J., Yakub,S., Yen,J., Yoon,L., Yoon,V.,
Yu,F., Zhang,J., Zhou,X., Zhou,X., Zhao,S., Dunn,D., von
Wiederhausen,A., Weiss,R., Smith,D.R., Holt,R.A., Smith,H.O.,
Weinstock,G. and Gibbs,R.A.
Direct Submission
Unpublished
2 (bases 1 to 215223)
Worley,K.C.
Direct Submission
Submitted (19-JUL-2002) Human Genome Sequencing Center, Department
of Molecular and Human Genetics, Baylor College of Medicine, One
Baylor Plaza, Houston, TX 77030, USA
3 (bases 1 to 215223)
Rat Genome Sequencing Consortium.
Direct Submission
Submitted (12-OCT-2002) Human Genome Sequencing Center, Department
of Molecular and Human Genetics, Baylor College of Medicine, One
Baylor Plaza, Houston, TX 77030, USA
On Oct 12, 2002 this sequence version replaced gi:21908474.
The sequence in this assembly is a combination of BAC based reads
and whole genome shotgun sequencing reads assembled using Atlas
(http://www.hgsc.bcm.tmc.edu/projects/rat/). Each contig described
in the feature table below represents a scaffold in the Atlas
assembly (a 'contig-scaffold'). Within each contig-scaffold,
individual sequence contigs are ordered and oriented, and separated
by sized gaps filled with Ns to the estimated size. The sequence
may extend beyond the ends of the clone and there may be sequence
contigs within a contig-scaffold that consist entirely of whole
genome shotgun sequence reads. Both end sequences and whole genome
shotgun sequence only contigs will be indicated in the feature
table.
----- Genome Center
Center: Baylor College of Medicine
Center code: BCM
Web site: http://www.hgsc.bcm.tmc.edu/
Contact: hgsc-help@bcm.tmc.edu
----- Project Information
Center project name: GXSJ
Center clone name: CH230-142L1
----- Summary Statistics
Assembly program: Phrap; version 0.950329
Consensus quality: 200801 bases at least Q40
Consensus quality: 203059 bases at least Q30
Consensus quality: 204719 bases at least Q20
Estimated insert size: 206878; sum-of-contigs estimation
Quality coverage: 7x in Q20 bases; sum-of-contigs estimation

* NOTE: Estimated insert size may differ from sequence length
* (see http://www.hgsc.bcm.tmc.edu/docs/Genbank_draft_data.html).
* NOTE: This is a 'working draft' sequence. It currently
* consists of 4 contigs. The true order of the pieces
* is not known and their order in this sequence record is

- * arbitrary. Gaps between the contigs are represented as
- * runs of N, but the exact sizes of the gaps are unknown.
- * This record will be updated with the finished sequence
- * as soon as it is available and the accession number will
- * be preserved.

*	1	28223:	contig of 28323 bp in length
*		28223:	gap of unknown length
*	28324	28423:	contig of 27205 bp in length
*	28424	55228:	gap of unknown length
*	55729	213935:	contig of 158207 bp in length
*	55729	213935:	gap of unknown length
*	213936	214035:	contig of 1188 bp in length.
*	214036	215223:	contig of 1188 bp in length.

```

FEATURES
  source
    Location/Qualifiers
      1. .215223

```

```

source
misc_feature
/clone="CH230-14211"
/db_xref="taxon:10116"
/mol_type="genomic DNA"
/organism="Rattus norvegicus"
1. .219223
clone_end:77
/note="clone_boundary"
complement(6631..6964)

```

site: ECORT

size:ECOR1
end_sequence:RWAXO61TJB"
28324. .28423
gap

```

misc feature
/estimated_length=unknown
28424. 29841

```

- /note="wgs_contig"

```
gap
55629.55728
/estimated_length=unknown
```

```
misc feature 55729. .58010
```

```

gap
/|note="wgs_contig"
213936. 214035

```

```

/estimated_length=unknown

```

ORIGIN

Query Match	20.1%;	Score 140.6;	DB 14;	Length 215223;
Best Local Similarity	65.0%;	Pred. No. 2.8e-27;		
Matches 236;	Conservative	3;	Mismatches 117;	Indels 7; Gaps 2;
Qy	53	AAAAAATGCTTTCAATAAATATATGATAAAGGACTTATATTTTTCAAGCCATAGAGTC	112	
Db	25680	AAAAGTTGCTTTCAATAAATATATGCTGATAGAGATCATAGATCTCTGAGCTCTGGGGCC	25621	
Qy	113	ATTTCCTCTGAAGCATCTTGGCGAAGTCATCCCACTGTTCCTGAGAGTGGCGAGTGA	172	
Db	25620	ATTAGCTCTGGAGCATGTGTTTAAGTTAGCTTCGAATGTTGAAAAAGCCCGTAGGTGA	25561	
Qy	173	GGGCTGACCTATTGCTCTGCACCTTACTCTCTATCTCAGCTGTCCCTCCCACTTTTCCAGGTG	232	
Db	25560	GAACCGACTTCCGGTGGTCCAC-----GCCTCTCTTCGGCTCTCTCTTCGCGCAGGTG	25506	
Qy	233	CTGCCAGACACATGACAACTGTGTAYGACGAGGCCAAGAAGCTGGACAGCTGTAAATTTCT	292	
Db	25505	CTGCCAGACTCATGACCCACTGTCTACAATCAGGCCAAGAAGCTGGAAAGCTGTAAATTCCT	25446	
Qy	293	GCTGGACAMMCCGTACACCCACCACTTATTCATCTCGTCTGCTCTGGCTGGCAATCACCTG	352	
Db	25445	CATCGACAACCCCTTACACCAACACGACTCATACAAGTGCTTCGGGAACGTGATCACCTG	25386	
Qy	353	TAGCAGTAGGTTTATCCCTTCTTGACCTATGAATTTCTAGTTGGTTTCTCAGTAGGCCGGG	412	
Db	25385	CAGCGGTAGGGTTA--CCCTCCGGGACCTCTGGACTCTGCCTGGCCTTTGGCCGGCTCTGG	25328	
Qy	413	GGG 415		
Db	25327	AGG 25325		

RESULT 10
AC097575
LOCUS
AC097575 23861 bp DNA linear HTG 13-MAY-2003
Rattus norvegicus clone CH230-20A7, WORKING DRAFT SEQUENCE, 4
DEFINITION
unordered pieces.

ACCESSION
VERSION
KEYWORDS
SOURCE
ORGANISM

REFERENCE AUTHORS

TITLE	JOURNAL	REFERENCE	AUTHORS	TITLE	JOURNAL
1. The Effect of Temperature on the Rate of Reaction of Hydrogen Peroxide with Potassium Iodide	Journal of Chemical Education	1965	Smith, J. D.	2. The Effect of Concentration on the Rate of Reaction of Hydrogen Peroxide with Potassium Iodide	Journal of Chemical Education
3. The Effect of pH on the Rate of Reaction of Hydrogen Peroxide with Potassium Iodide	Journal of Chemical Education	1965	Smith, J. D.	4. The Effect of Catalyst on the Rate of Reaction of Hydrogen Peroxide with Potassium Iodide	Journal of Chemical Education
5. The Effect of Surface Area on the Rate of Reaction of Hydrogen Peroxide with Potassium Iodide	Journal of Chemical Education	1965	Smith, J. D.	6. The Effect of Temperature on the Rate of Reaction of Hydrogen Peroxide with Potassium Iodide	Journal of Chemical Education

REFERENCE
AUTHORS
TITLE
JOURNAL

COMMENT

AC0957575
AC0957575_6 GI:30581004
HTG: HTGS PHASE1: HTGS DRAFT: HTGS_FULLTOP.
Rattus norvegicus (Norway rat)
Rattus norvegicus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Euarchontoglires; Glires; Rodentia;
Sciuridae; Muridae; Murinae; Rattus.

1 (bases 1 to 238961)

Muzny, D. Marie., Metzker, M. Lee., Abramson, S., Adams, C., Alder, J., Allen, C., Allen, H., Alsbrooks, S., Amin, A., Anguiano, D., Anyalebechi, V., Aoyagi, A., Ayodeji, M., Baca, S., Baden, H., Baldwin, D., Bandaranaike, D., Barber, M., Barnstead, M., Benahmed, F., Biswal, K., Blair, J., Blankenburg, K., Blyth, P., Brown, M., Bryant, N., Buhay, C., Burch, P., Burrell, K., Calderon, E., Cardenas, V., Carter, K., Cavazos, I., Ceasar, H., Center, A., Chacko, J., Chavez, D., Chen, G., Chen, R., Chen, Y., Chen, Z., Chu, J., Cleveland, C., Cockrell, R., Cox, C., Coyle, M., Cree, A., D'Souza, L., Davila, M. L., Davis, C., Davy-Carroll, L., De Anda, C., Dederich, D., Delgado, O., Denson, S., Deramco, C., Ding, Y., Dinh, H., Diviya, K., Draper, H., Dugan-Rocha, S., Dunn, A., Durbin, K., Duval, B., Eaves, K., Egan, A., Escotto, M., Eugene, C., Evans, C. A., Falls, T., Fan, G., Fernandez, S., Finley, M., Flagg, N., Forbes, L., Foster, M., Foster, P., Fraser, C. M., Gabisi, A., Gatica, R., Garcia, A., Garner, T., Garza, M., Gebregeorgis, E., Geer, K., Gill, R., Grady, M., Guerra, W., Guevara, W., Gunaratne, P., Haaland, W., Hamil, C., Hamilton, C., Hamilton, K., Harvey, Y., Havlak, P., Hawes, A., Henderson, N., Hernandez, J., Hernandez, R., Hines, S., Hladun, S. L., Hodgson, A., Hogues, M., Hollins, B., Howells, S., Hulvik, S., Hume, J., Idlebird, D., Jackson, A., Jackson, L., Jacob, L., Jiang, H., Johnson, B., Johnson, R., Jolivet, A., Karpathy, S., Kelly, S., Kelly, S., Khan, Z., King, L., Kovar, C., Kowis, C., Kraft, C. L., Lebow, H., Levay, J., Lewis, L., Li, Z., Liu, J., Liu, J., Liu, W., Liu, Y., London, P., Longacre, S., Lopez, J., Lorensuewa, L., Loulseg, H., Lozardo, R. J., Lu, X., Ma, J., Maheshwari, M., Mahindartine, M., Mahmoud, M., Malloy, K., Mangum, B., Mapua, P., Martin, K., Martin, R., Meenen, E., Mathwney, S., McLeod, M. P., McNeill, T. Z., Montemayor, J., Moore, S., Milosavljevic, A., Miner, G., Minja, E., Montemada, M., Murphy, M., Naïr, L., Morgan, M., Morris, K., Morris, S., Munidasa, M., Nguyen, N., Norris, S., Nankervis, C., Neal, D., Newton, N., Nguyen, N., Parks, K., Nwaokeme, O., Okunonu, G., Olarnpunsagoon, A., Pal, S., Parks, K., Pasternak, S., Paul, H., Perez, A., Perez, L., pfannkoch, C., Pioppfer, F., Poindexter, A., Popovic, D., Primus, E., Pu, L., Pua, M., Quiroz, J., Rachlin, E., Reeves, K., Regier, M. A., Reigh, R., Reilly, B., Reilly, M., Ren, Y., Reuter, M., Richards, S., Riggs, F., Rives, C., Rodkey, T., Rojas, A., Rose, M., Rose, R., Ruiz, S. J., Sanders, W., Savary, G., Scherer, S., Scott, G., Shatsman, S., Shen, H., Shetty, J., Shivartsbeyn, A., Sisson, I., Sitter, C. D., Smajs, D., Sneed, A., Sodergren, E., Song, A.-Z., Soreller, R., Sosa, J., Steimle, M., Strong, R., Sutton, A., Svatek, A., Tajos, Z., Taylor, C., Taylor, T., Thomas, N., Thomas, S., Tingey, A., Trejos, P., Usmani, K., Valas, R., Vera, V., Villalana, D., Waldron, L., Walker, B., Wang, J., Wang, Q., Wang, S., Warren, J., Warren, R., Wei, X., White, F., Williams, G., Willson, R., Wleczyk, R., Wooden, H., Worley, K., Wright, D., Wright, R., Wu, J., Yakub, S., Yen, J., Yoon, L., Yoon, V., Yu, F., Zhang, J., Zhou, J., Zhou, X., Zhao, S., Dunn, D., von Niederhausern, A., Weiss, R., Smith, D. R., Holt, R. A., Smith, H. O., Weinstock, G. and Gibbs, R. A.

Direct Submission
Unpublished
2 (bases 1 to 238961)
Worley, K.C.
Direct Submission
Submitted (19-OCT-2001) Human Genome Sequencing Center, Department
of Molecular and Human Genetics, Baylor College of Medicine, One
Baylor Plaza, Houston, TX 77030, USA
3 (bases 1 to 238961)
Rat Genome Sequencing Consortium.
Direct Submission
Submitted (13-MAY-2003) Human Genome Sequencing Center, Department
of Molecular and Human Genetics, Baylor College of Medicine, One
Baylor Plaza, Houston, TX 77030, USA
On May 13, 2003 this sequence version replaced gi:23096385.

The sequence in this assembly is a combination of BAC based reads and whole genome shotgun sequencing reads assembled using Atlas (<http://www.hgsc.bcm.tmc.edu/projects/rat/>). Each contig described in the feature table below represents a scaffold in the Atlas assembly (a 'contig-scaffold'). Within each contig-scaffold, individual sequence contigs are ordered and oriented, and separated by sized gaps filled with Ns to the estimated size. The sequence may extend beyond the ends of the clone and there may be sequence contigs within a contig-scaffold that consist entirely of whole genome shotgun sequence reads. Both end sequences and whole genome shotgun sequence only contigs will be indicated in the feature table.

----- Genome Center

Center: Baylor College of Medicine
Center code: BCM
Web site: <http://www.hgsc.bcm.tmc.edu/>
Contact: hgsc-help@bcm.tmc.edu

----- Project Information

Center project name: GIUS
Center clone name: CH230-20A7
----- Summary Statistics

Assembly program: Atlas 3.0:

Consensus quality: 222002 bases at least Q40

Consensus quality: 225364 bases at least Q30

Consensus quality: 227894 bases at least Q20

Estimated insert size: 232967; sum-of-contigs estimation

Quality coverage: 7x in Q20 bases; sum-of-contigs estimation

* NOTE: Estimated insert size may differ from sequence length (see http://www.hgsc.bcm.tmc.edu/docs/Genbank_draft_data.html).
* NOTE: This is a 'working draft' sequence. It currently consists of 4 contigs. The true order of the pieces is not known and their order in this sequence record is arbitrary. Gaps between the contigs are represented as runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence as soon as it is available and the accession number will be preserved.

* 1 234563: contig of 234563 bp in length
* 234564 234663: gap of unknown length
* 234664 236006: contig of 1343 bp in length
* 236007 236106: gap of unknown length
* 236107 237379: contig of 1273 bp in length
* 237380 237479: gap of unknown length
* 237480 238961: contig of 1482 bp in length.

FEATURES

source

1. .238961
/organism="Rattus norvegicus"
/mol_type="genomic DNA"
/db_xref="taxon:10116"
/clone="CH230-20A7"

misc_feature

1. .1200
/note="wgs_end_extension
clone_end:Sp6"
15551. .15691
/note="clone boundary
clone_end:Sp6"

misc_feature

site:EcoRI
end_sequence:BH317687"
21638. .67975
/note="clone boundary
clone_end:T7"

misc_feature

site:EcoRI
end_sequence:BH317686"
79112. .81297
/note="wgs_end_extension
clone_end:T7"

misc_feature

233264. .234563
/note="wgs_end_extension
clone_end:T7"
234564. .234663
/estimated_length=unknown

gap

236007. .236106

gap

gap .. /estimated_length=unknown
237380. .237479
/estimated_length=unknown
ORIGIN
Query Match 20.1%; Score 140.6; DB 14; Length 238961;
Best Local Similarity 65.0%; Pred. No. 2.8e-27;
Matches 236; Conservative 3; Mismatches 117; Indels 7; Gaps 2;
QY 53 AAAAAATCCTTTCAATAATATATCAATAAGGACTATATTTTCAAGCCATAGGATC 112
DB 73855 AAAAGTTGCTTTCAATAAATCTATGCTGATAGATCATAGATCCTCGAGCC 73914
QY 113 ATTTCTCTCGAAGCATCTTGGCAAGTCATCCCACTGTTCTCTGAGAGTGGGAGTGA 172
DB 73915 ATTAGCTCTGGAGCACTGTGTTAAGTTAGCTGAATGTTGAAAGCCGTAGTGA 73974
QY 173 GGGGTGACCTATGCTCTGCACTTACTCTCTATCTCAGCTGCTCCCTCCCACTTTCCAGTG 232
DB 73975 GAACCGACTTCCGGTGGTCCAC-----GCCTCTCTTCCGCTCTTCTCTCGCCAGTG 74029
QY 233 CTGCCAGACATGACACTCTGTAYGACCGAGCCAGAGCTGCACAGCTTAATTTCT 292
DB 74030 CTGCCAGACTCATGCCACTCTCTCAATCAGGCCAAGAAGCTGGAAGCTGTAAATTCCT 74089
QY 293 GCTGCACAMMCCGTACACCCACACTATTCTACTCTGCTGCTCTGCTCGCAATCACCTG 352
DB 74090 CATGCANACCCCTACACCAACAGTACTCATAGTCTCCGGAAAGCTGATCACCTG 74149
QY 353 TAGCAGTAGGTTTATCCCTTCTTGACCTATGAATCTAGTTGTTCTTCAGTAGCCCGG 412
DB 74150 CAGCGTAGGGTTA--CCCTCCGGGACCTCTGGACTCTGCGCTTGGCGCTTGGCGGCTCGG 74207
QY 413 GGG 415
DB 74208 AGG 74210

RESULT 11

AC096966/c

LOCUS

DEFINITION

***, 12 unordered pieces.

AC096966

VERSION

AC096966.6 GI:30521249

KEYWORDS

HTG; HTGS_PHASE1; HTGS_DRAFT; HTGS_ENRICHED.

SOURCE

Rattus norvegicus (Norway rat)

ORGANISM

Rattus norvegicus

Eukaryota; Metazoa;

Mammalia; Eutheria; Euarchontoglires; Glires; Rodentia;

Sciurognathi; Muroidae; Muridae; Rattus;

1 (bases 1 to 290053)

REFERENCE

AUTHORS

Muzny,D,Marie., Metzker,M, Lee., Abramson,S., Adams,C., Alder,J.,

Allen,C., Allen,H., Alsbrooks,S., Amin,A., Anguiano,D.,

Anyalebechi,V., Aoyagi,A., Ayodeji,M., Baca,E., Baden,H.,

Baldwin,D., Bandaranaike,D., Barber,M., Barnstead,M., Benahmed,F.,

Biswal,K., Blair,J., Blankenburg,K., Blyth,P., Brown,M.,

Bryant,N., Buhay,C., Burch,P., Burrell,K., Calderon,E.,

Cardenas,V., Carter,K., Cavazos,I., Ceasar,H., Center,A.,

Chacko,J., Chavez,D., Chen,G., Chen,R., Chen,Y., Chen,Z., Chu,J.,

Cleveland,C., Cockrell,R., Cox,C., Coyle,M., Cree,A., D'Souza,L.,

Davila,M.L., Davis,C., Davy-Carroll,L., De Anda,C., Dederich,D.,

Delgado,O., Denson,S., Deramo,C., Ding,Y., Dinh,H., Divya,K.,

Draper,H., Dugan-Rocha,S., Dunn,A., Durbin,K., Duval,B., Eaves,K.,

Egan,A., Escotto,M., Eugene,C., Evans,C.A., Fall,T., Fan,G.,

Fernandez,S., Finley,M., Flagg,N., Forbes,L., Foster,M., Foster,P.,

Fraser,C.M., Gabisi,A., Ganta,R., Garcia,A., Garner,T., Garza,M.,

Georgiadis,E., Geer,K., Gill,R., Grady,M., Guerra,W., Guvavara,W.,

Harvey,Y., Havlak,P., Hawes,A., Henderson,N., Hernandez,J.,

Hernandez,R., Hines,S., Hladun,S.L., Hodgson,A., Hogues,M.,

Hollins,B., Howells,S., Hulyk,S., Hume,J., Idlebird,D., Jackson,A.,

Jackson,L., Jacob,L., Jiang,H., Johnson,B., Johnson,R., Jolivet,A.,

Karpathy, S., Kelly, S., Kelly, S., Khan, Z., King, L., Kovar, C.,
Kowis, C., Kraft, C.L., Lebow, H., Levan, J., Lewis, L., Li, Z., Liu, J.,
Liu, J., Liu, W., Liu, Y., London, P., Longacre, S., Lopez, J.,
Lorensuwa, L., Louised, H., Lozado, R.J., Lu, X., Ma, J.,
Maheshwari, M., Mahindartne, M., Mahmood, M., Malloy, K., Mangum, A.,
Mangum, B., Mapua, P., Martin, K., Martin, R., Martinez, E.,
Mawhney, S., McLeod, M.P., McNeill, T.Z., Meenen, E.,
Milosavljevic, A., Miner, G., Minja, E., Montemayor, J., Moore, S.,
Morgan, M., Morris, K., Morris, S., Munidasa, M., Murphy, M., Nair, L.,
Nankervis, C., Neal, D., Newton, N., Nguyen, N., Norris, S.,
Nwaokemele, O., Okwuonu, G., Olarnpunsagoon, A., Pal, S., Parks, K.,
Pasternak, S., Paul, H., Perez, A., Perez, L., Pfankuch, C.,
Plopper, F., Poindexter, A., Popovic, D., Primus, E., Pu, L.-L.,
Puar, M., Quiroz, J., Rachlin, E., Reeves, K., Regier, M.A., Reigh, R.,
Reilly, B., Reilly, M., Ren, Y., Reuter, M., Richards, S., Riggs, F.,
Rives, C., Rodkey, T., Rojas, A., Rose, M., Rose, J., Ruiz, S.J.,
Sanders, W., Savary, G., Scherer, S., Scott, G., Shatsman, S., Shen, H.,
Shetty, J., Shvartsbeyn, A., Sison, I., Sitter, C.D., Smajs, D.,
Sneed, A., Sodergren, E., Song, X.-Z., Sorelle, R., Sosa, J., Taylor, C.,
Steinle, M., Strong, R., Sutton, A., Svatek, A., Tabor, P., Taylor, C.,
Taylor, T., Thomas, N., Thomas, S., Tingey, A., Trejos, Z., Umani, K.,
Valas, R., Vera, V., Villasan, D., Waldron, L., Walker, B., Wang, J.,
Wang, Q., Wang, S., Warren, J., Warren, R., Wei, X., White, F.,
Williams, G., Willson, R., Wleczek, R., Wooden, H., Worley, K.,
Wright, D., Wright, R., Wu, J., Yakub, S., Yen, J., Yoon, L., Yoon, V.,
Yu, F., Zhang, J., Zhou, J., Zhou, X., Zhao, S., Dunn, D., von
Niederhausern, A., Weiss, R., Smith, D.R., Holt, R.A., Smith, H.O.,
Weinstock, G. and Gibbs, R.A.

TITLE JOURNAL

REFERENCE

AUTHORS

TITLE

JOURNAL

REFERENCE

AUTHORS

TITLE

JOURNAL

COMMENT

FEATURES
source

1. .290053

/organism="Rattus norvegicus"

/mol_type="genomic DNA"

/db_xref="taxon:10116"

/clone="CH230-213K13"

27648. 27747

/estimated_length=unknown

35097. .36658

/note="wgs_contig"

44240. .44339

/estimated_length=unknown

44340. .45348

/note="wgs_contig"

270030. .271309

/note="wgs_contig"

271310. .271409

/estimated_length=unknown

272831. .272930

/estimated_length=unknown

274122. .274221

/estimated_length=unknown

275625. .275724

/estimated_length=unknown

277024. .277123

/estimated_length=unknown

278864. .278963

/estimated_length=unknown

280727. .280826

/estimated_length=unknown

284234. .284333

/estimated_length=unknown

286732. .286831

/estimated_length=unknown

ORIGIN

Query Match 20.1%; Score 140.6; DB 14; Length 290053;

Best Local Similarity 65.0%; Pred. No. 2.8e-27; Indels 7; Gaps 2;

Matches 236; Conservative 3; Mismatches 117;

QY 53 AAAAAATGCTTCAATAAATATATGATAAAAGACCTTATATATTTTCAAGCCATAGATC 112

Assembly program: Atlas 3.0;

Consensus quality: 249124 bases at least Q40

Consensus quality: 252995 bases at least Q30

Consensus quality: 256136 bases at least Q20

Estimated insert size: 261662; sum-of-contigs estimation

Quality coverage: 6x in Q20 bases; sum-of-contigs estimation

* NOTE: Estimated insert size may differ from sequence length

(see http://www.hgsc.bcm.tmc.edu/docs/Genbank_draft_data.html)

```
Db 65775 AAAAGTGTCTTTCAATAAATCTATGCTGATAGAGATCATAGATCTCGAGCCTCGGGGCC 65716
Qy 113 ATTTCCTCTGAAGCATCTTGGGAAAGTATCCACCTGCTCTGAGAGTGGCGAGGTGA 172
Db 65715 ATTAGCTCTGGAGCACTGTGTTAAGTAGCTCGAATGTTTGAAGAGCCCGTAGGTGA 65656
Qy 173 GGGCTGACCTATTGCTCTGCATCTTACTCTATCTAGCTGTGCTCCCTCCCACTTTCAGGTG 232
Db 65655 GAACGACTTCGGTGGTCCAC-----GCCTCTCCTTCGGCTCTTCTCTCGCCAGGTG 65601
Qy 233 CTGCCAGACATGACAAGCTGTAYGACCCAGCGGAGAGCTGGACAGCTGTAATTTCT 292
Db 65600 CTGCCAGACTCAAGACCACTGTCTACAAATCAGCCCAAGAGCTGGAAGCTGTAAATTCCT 65541
Qy 293 GCTGGACAMMCGGTACACCCACACCTATTATCTACTCTGCTCTGGCTGGCAATCACCTG 352
Db 65540 CATCGACACCCCTACACCAACACGTAATCTACAGTGTCTCGGAGAGCTGATCACCTG 65481
Qy 353 TAGCAGTAGGTTATCCCTTCCTTGACCTATGAATTTCTAGTTGTTCTCAGTAGCCGGG 412
Db 65480 CAGCGTAGGGTTA--CCCTCCGGGACCTCTGGACTCTGCTGCGCTTGGCGCGTCTGG 65423
Qy 413 GGG 415
Db 65422 AGG 65420

RESULT 12
CQ922146
LOCUS CQ922146 418 bp DNA linear PAT 23-NOV-2004
DEFINITION Sequence 3346 from Patent WO2004097052.
ACCESSION CQ922146
VERSION CQ922146.1 GI:56212087
KEYWORDS Homo sapiens (human)
SOURCE Homo sapiens
ORGANISM Homo sapiens
REFERENCE 1
AUTHORS Burczynski, M.E., Twine, N.C., Slonim, D.K., Trepicchio, W.L.,
Strahs, A., Immerman, F. and Dorner, A.J.
TITLE Methods for prognosis and treatment of solid tumors
JOURNAL Patent: WO 2004097052-A 3346 11-NOV-2004;
Wyeth (US); Burczynski, Michael E. (US)
FEATURES
source
1..418
/organism="Homo sapiens"
/mol_type="unassigned DNA"
/db_xref="taxon:9606"
ORIGIN
Query Match 18.7%; Score 130.8; DB 6; Length 418;
Best Local Similarity 94.3%; Pred. No. 1.5e-24;
Matches 132; Conservative 3; Mismatches 5; Indels 0; Gaps 0;
Qy 221 ACTTTCCAGGTGCTCCAGACACATGACAACCTGTCTAYGACCCAGCGGAGAGCTGGACAG 280
Db 185 ACTGACAAGTGTCTCCAGACACATGACAACCTGTCTAYGACCCAGCGGAGAGCTGGACAG 244
Qy 281 CTGTAATTTCTGCTGGACAMMCCGTACACCCACACCTATTATCTACTCTGCTCTGGCTC 340
Db 245 CTGTAATTTCTGCTGGACAMMCCGTACACCCACACCTATTATCTACTCTGCTCTGGCTC 304
Qy 341 GGCAATCACTGTAGCAGTA 360
Db 305 GGCAATCACTGTAGCAGTA 324

RESULT 13
AX377240
LOCUS AX377240 447 bp DNA linear PAT 18-MAR-2002
```

```
DEFINITION Sequence 2 from Patent WO0212562.
ACCESSION AX377240
VERSION AX377240.1 GI:19573529
KEYWORDS Homo sapiens (human)
SOURCE Homo sapiens
ORGANISM Homo sapiens
REFERENCE 1
AUTHORS Kazemi, A., Kliem, S.E. and Koshy, B.
TITLE Haplotypes of the pla2glb gene
JOURNAL Patent: WO 0212562-A 2 14-FEB-2002;
Genaisance Pharmaceuticals, Inc. (US)
FEATURES
source
1..447
/organism="Homo sapiens"
/mol_type="unassigned DNA"
/db_xref="taxon:9606"
ORIGIN
Query Match 18.7%; Score 130.8; DB 6; Length 447;
Best Local Similarity 94.3%; Pred. No. 1.5e-24;
Matches 132; Conservative 3; Mismatches 5; Indels 0; Gaps 0;
Qy 221 ACTTTCCAGGTGCTCCAGACACATGACAACCTGTCTAYGACCCAGCGGAGAGCTGGACAG 280
Db 186 ACTGACAAGTGTCTCCAGACACATGACAACCTGTCTAYGACCCAGCGGAGAGCTGGACAG 245
Qy 281 CTGTAATTTCTGCTGGACAMMCCGTACACCCACACCTATTATCTACTCTGCTCTGGCTC 340
Db 246 CTGTAATTTCTGCTGGACAMMCCGTACACCCACACCTATTATCTACTCTGCTCTGGCTC 305
Qy 341 GGCAATCACTGTAGCAGTA 360
Db 306 GGCAATCACTGTAGCAGTA 325

RESULT 14
CQ723247
LOCUS CQ723247 559 bp DNA linear PAT 03-FEB-2004
DEFINITION Sequence 9181 from Patent WO02068579.
ACCESSION CQ723247
VERSION CQ723247.1 GI:42284104
KEYWORDS Homo sapiens (human)
SOURCE Homo sapiens
ORGANISM Homo sapiens
REFERENCE 1
AUTHORS Venter, C.J., Adams, M.C., Li, P.W. and Myers, E.W.
TITLE Kits, such as nucleic acid arrays, comprising a majority of
humanexons or transcripts, for detecting expression and other uses
thereof
JOURNAL Patent: WO 02068579-A 9181 06-SEP-2002;
PE Corporation (NY) (US)
FEATURES
source
1..559
/organism="Homo sapiens"
/mol_type="unassigned DNA"
/db_xref="taxon:9606"
ORIGIN
Query Match 18.7%; Score 130.8; DB 6; Length 559;
Best Local Similarity 94.3%; Pred. No. 1.5e-24;
Matches 132; Conservative 3; Mismatches 5; Indels 0; Gaps 0;
Qy 221 ACTTTCCAGGTGCTCCAGACACATGACAACCTGTCTAYGACCCAGCGGAGAGCTGGACAG 280
Db 221 ACTGGACAAGTGTCTGGACACATGACAACCTGTCTATGACCCAGCGGAGAGCTGGACAG 280
Qy 281 CTGTAATTTCTGCTGGACAMMCCGTACACCCACACCTATTATCTACTCTGCTCTGGCTC 340
```

Db

281

CTGTAAATTTCTGCTGGACAACCCGTACACCCACACCTATTTCATCTCGTGTCTGGCTC

340

Qy

341

GGCAATCACCTGTAGCAGTA

360

Db

341

GGCAATCACCTGTAGCAGCA

360

RESULT 15

AR380655

LOCUS

AR380655

562 bp

DNA

Linear

PAT 18-DEC-2003

DEFINITION

Sequence 1200 from patent US 6607879.

ACCESSION

AR380655

VERSION

AR380655.1

GI:40088289

KEYWORDS

Unknown.

SOURCE

Unknown.

ORGANISM

Unknown.

REFERENCE

1 (bases 1 to 562)

AUTHORS

Cocks, B.G., Stuart, S.G. and Seilhamer, J.J.

TITLE

Compositions for the detection of blood cell and immunological response gene expression

JOURNAL

Patent: US 6607879-A 1200 19-AUG-2003;

FEATURES

Incyte Corporation; Palo Alto, CA

location/Qualifiers

1..562

source

/organism="unknown"

/mol_type="genomic DNA"

ORIGIN

Query Match

18.7%

Score 130.8;

DB 6;

Length 562;

Best Local Similarity

94.3%;

Pred. No. 1.5e-24;

Matches 132;

Conservative 3;

Mismatches 5;

Indels 0;

Gaps 0;

Qy

221

ACTTTCCAGGTGTCGACACATGACAACCTGCTAYGACCAAGCCAGCAAGCTGGACAG

280

Db

223

ACTGGACAAGTGTGCCAGACATGACAACCTGCTATGACCAAGCCAGCAAGCTGGACAG

282

Qy

281

CTGTAAATTTCTGCTGGACAMCCGTACACCCACACCTATTTCATCTCGTGTCTGGCTC

340

Db

283

CTGTAAATTTCTGCTGGACACCCGTACACCCACACCTATTTCATCTCGTGTCTGGCTC

342

Qy

341

GGCAATCACCTGTAGCAGTA

360

Db

343

GGCAATCACCTGTAGCAGCA

362

Search completed: February 9, 2006, 20:42:23
Job time : 3498.02 secs

GenCore version 5.1.7
Copyright (c) 1993 - 2006 Bioacceleration Ltd.

OM nucleic - nucleic search, using sw model

Run on: February 9, 2006, 14:19:56 ; Search time 451.558 Seconds
(without alignments)
10346.286 Million cell updates/sec

Title: US-10-607-806-1-G7328_COPY_7000_7700

Perfect score: 699.8

Sequence: 1 gtctgtcactgtgtccag.....Tgcagctcaacctctctgag 701

Scoring table: IDENTITY NUC

Gapop 10_0 , Gapext 1.0

Searched: 4996997 seqs, 3332346308 residues

Total number of hits satisfying chosen parameters: 9993994

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : N_Geneseq_21.*

- 1: Geneseqn1980s.*
- 2: Geneseqn1990s.*
- 3: Geneseqn2000s.*
- 4: Geneseqn2001as.*
- 5: Geneseqn2001bs.*
- 6: Geneseqn2002as.*
- 7: Geneseqn2002bs.*
- 8: Geneseqn2003as.*
- 9: Geneseqn2003bs.*
- 10: Geneseqn2003cs.*
- 11: Geneseqn2003ds.*
- 12: Geneseqn2004as.*
- 13: Geneseqn2004bs.*
- 14: Geneseqn2005s.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	699.8	100.0	12174	12 ADJ09983	Adj09983 Human pho
2	699.8	100.0	13612	6 ABK47376	Abk47376 Human pho
3	699.4	100.0	12174	12 ADI35082	Adi35082 Human PLA
4	380.8	54.4	3375	3 AAA35101	Aaa35101 Human ade
5	380.8	54.4	3375	3 AAF21223	Aaf21223 Human low
6	380.8	54.4	3375	10 ABD20766	Abd20766 Human nuc
7	380.8	54.4	3375	11 ABD20766	Abd20766 Human pul
8	380.8	54.4	6771	3 AAA35103	Aaa35103 Human ade
9	380.8	54.4	6771	3 AAF21225	Aaf21225 Human low
10	380.8	54.4	6771	10 ABD20768	Abd20768 Human nuc
11	380.8	54.4	6771	11 ABD20768	Abd20768 Human pul
12	130.8	18.7	418	13 ADU12907	Adu12907 Solid tum
13	130.8	18.7	447	6 ABK47377	Abk47377 Human pho
14	130.8	18.7	447	14 ADV43454	Adv43454 Human psy
15	130.8	18.7	562	3 AAA35100	Aaa35100 Human ade
16	130.8	18.7	562	3 AAF21222	Aaf21222 Human low
17	130.8	18.7	562	6 ABL64906	AbL64906 Lung canc
18	130.8	18.7	562	6 ABL65512	AbL65512 Lung canc
19	130.8	18.7	562	6 ABL65915	AbL65915 Lung canc

20	130.8	18.7	562	10 ABZ96916	Abz96916 Human nuc
21	130.8	18.7	562	11 ADI31874	Adi31874 Human cdn
22	130.8	18.7	562	11 ABD20765	Abd20765 Human pul
23	130.8	18.7	562	13 ADS83941	Ads83941 Human lym
24	130.8	18.7	585	12 ADO28598	Ado28598 Human PA2
25	130.8	18.7	630	3 AAC98824	Aac98824 Human pan
26	130.8	18.7	644	2 AAH41391	Aah41391 Human nor
27	130.8	18.7	654	4 AAH57532	Aah57532 Human pan
28	130.8	18.7	654	5 AAS76099	Aas76099 DNA encod
29	130.8	18.7	742	13 ACN42348	Acn42348 Human dia
30	126.2	18.0	496	4 AAH57324	Aah57324 Human pan
31	126	18.0	26593	13 ABD33464	Abd33464 Murine ca
32	121.8	17.4	445	3 AAC99186	Aac99186 Human pan
33	89.8	12.8	369	1 AAN81342	Aan81342 DNA encod
34	86.6	12.4	223556	11 ACN44110	Acn44110 Human gen
35	86.6	12.4	243428	12 ADP51132	Adp51132 Human P-R
36	85.6	12.2	96596	9 ADA02834	Ada02834 Human MEF
37	85.6	12.2	96596	10 ADB72572	Adb72572 Human MEF
38	85.6	12.2	96596	10 ADC85313	Adc85313 Mouse Mef
39	85.6	12.2	96596	12 ADM74429	Adm74429 Human car
40	85	12.1	460	2 AAQ04439	Aaq04439 Gene enco
41	83.4	11.9	460	1 AAN91424	Aan91424 DNA encod
42	81.6	11.7	34551	10 ADC87424	Adc87424 Human GPC
43	81	11.6	3509	11 ADM02028	Adm02028 Human cdn
44	80	11.4	62658	13 ABD33339	Abd33339 Human can
45	80	11.4	143306	6 ABK49586	Abk49586 Human tra

ALIGNMENTS

RESULT 1

ADJ09983

ID ADJ09983 standard; DNA; 12174 BP.

AC ADJ09983;

XX 17-JUN-2004 (first entry)

XX Human phospholipase A2 (PLA2G1B) DNA SeqID 1.

XX human; gene; ds; fat reduction; fat deposition; phospholipase A2;

PLA2G1B; chromosome 12q24; single nucleotide polymorphism; SNP;

appetite suppressant; lipase inhibitor; exercise regimen; obesity;

non-insulin dependent diabetes mellitus; NIDDM; cardiovascular disorder;

hypertension; antidiabetic.

XX Homo sapiens.

XX Key

XX Location/Qualifiers

FT variation

FT /tag= a

FT /standard name= "Single nucleotide polymorphisms"

FT replace(839,a)

FT /tag= b

FT /standard name= "Single nucleotide polymorphisms"

FT replace(4050,a)

FT /tag= c

FT /standard name= "Single nucleotide polymorphisms"

FT replace(4689,t)

FT /tag= d

FT /standard name= "Single nucleotide polymorphisms"

FT replace(6282,a)

FT /tag= e

FT /standard name= "Single nucleotide polymorphisms"

FT replace(6358,c)

FT /tag= f

FT /standard name= "Single nucleotide polymorphisms"

FT replace(6653,t)

FT /tag= g

FT /standard name= "Single nucleotide polymorphisms"

FT replace(7256,t)

FT /tag= h

FT	variation	/standard_name= "single nucleotide polymorphisms"	
FT		replace(7300,a)	
FT		/*tag= i	
FT	variation	/standard_name= "Single nucleotide polymorphisms"	
FT		replace(7301,c)	
FT		/*tag= j	
FT	variation	/standard_name= "single nucleotide polymorphisms"	
FT		replace(7328,g)	
FT		/*tag= k	
FT	variation	/standard_name= "Single nucleotide polymorphisms"	
FT		replace(8062,c)	
FT		/*tag= l	
FT	variation	/standard_name= "single nucleotide polymorphisms"	
FT		replace(9182,t)	
FT		/*tag= m	
FT	variation	/standard_name= "Single nucleotide polymorphisms"	
FT		replace(11649,c)	
FT		/*tag= n	
FT		/standard_name= "single nucleotide polymorphisms"	
XX		WO2004002296-A2.	
PN			
XX		08-JAN-2004.	
XX		27-JUN-2003; 2003WO-US020831.	
XX		27-JUN-2002; 2002US-0392362P.	
XX		(SEQU-) SEQUENOM INC.	
PA			
XX		Adam GIR, Langdown ML, Deniseenko MF, Dennis E, Cantor C;	
PI		Rubin B;	
XX			
PI			
XX			
DR		WPI; 2004-071944/07.	
DR		P-PSDB; ADJ09984.	
XX			
XX		Identifying a candidate therapeutic for fat reduction, useful for	
PT		treating diabetes, by introducing a test molecule to a system comprising	
PT		PLA2G1B protein or nucleic acid, and determining the presence of	
PT		interaction between the compounds.	
XX			
PS		Claim 1; SEQ ID NO 1; 116pp; English.	
XX			
CC		This invention relates to a novel candidate therapeutic agent useful for	
CC		fat reduction and disorders related to fat depositions. Specifically, it	
CC		refers to polymorphic variations in the phospholipase A2 (PLA2G1B) DNA,	
CC		which is located on chromosome 12q24 and has been associated with central	
CC		fat deposition. The present invention describes methods to detect the	
CC		presence or absence of these single nucleotide polymorphisms of PLA2G1B,	
CC		in particular G7328A and T9182G, and subsequently provide treatment that	
CC		reduces fat deposition. This treatment may consist of an appetite	
CC		suppressant, a lipase inhibitor, a phospholipase inhibitor, an exercise	
CC		regimen, a dietary regimen, psychological counselling, psychotherapy or a	
CC		psychotherapeutic. Accordingly, PLA2G1B is a target for reducing fat	
CC		deposition and it can be used to treat both obesity and non-insulin	
CC		dependent diabetes mellitus (NIDDM), as well as cardiovascular disorders	
CC		such as hypertension. As such, it exhibits antidiabetic activity. This	
CC		polynucleotide sequence is the human PLA2G1B DNA of the invention.	
XX			
SQ		Sequence 12174 BP; 3220 A; 2996 C; 2739 G; 3219 T; 0 U; 0 Other;	
	Query Match	100.0%; Score 699.8; DB 12; Length 12174;	
	Best Local Similarity	99.6%; Pred. No. 4.4e-191;	
	Matches 698; Conservative	3; Mismatches 0; Indels 0; Gaps 0;	
Qy	1	GTGCGTGCACCTGCTCCAGCCTGGGTAACAGAGCAACTCTGTCTCAAAAAAAAATG 60	
Db	7000	GTGCGTGCACCTGCTCCAGCCTGGGTAACAGAGCAACTCTGTCTCAAAAAAAAATG 7059	
Qy	61	CTTTCATTAATATGATAAAGGACTTATATTTTCAAGCATAGGATCATTTCTCC 120	
Db	7060	CTTTCATTAATATGATAAAGGACTTATATTTTCAAGCATAGGATCATTTCTCC 7119	

Qy	121	TGAAGCATCTTGGCGAAGTCATCCCACTGTTTCTCTGAGTGGGAGGTGAGGGCTGAC 180	
Db	7120	TGAAGCATCTTGGCGAAGTCATCCCACTGTTTCTCTGAGTGGGAGGTGAGGGCTGAC 7179	
Qy	181	CTATTGCTCTGCACTTACTCCTATCTCAGCTGTCCTCCACACTTTCAGGTGCTGCCAGA 240	
Db	7180	CTATTGCTCTGCACTTACTCCTATCTCAGCTGTCCTCCACACTTTCAGGTGCTGCCAGA 7239	
Qy	241	CACATGACAACTGCTAYGACCGAGCCAAAGCTGGACAGCTGTAATTTCTGCTGACACA 300	
Db	7240	CACATGACAACTGCTATGACCGAGCCAAAGCTGGACAGCTGTAATTTCTGCTGACACA 7299	
Qy	301	MMCCGTACACCCACACCTATTATCTGCTCTGGCTCGGCAATCACCTGTAGCAGTA 360	
Db	7300	ACCGTACACCCACACCTATTATCTGCTCTGGCTCGGCAATCACCTGTAGCAGTA 7359	
Qy	361	GGTTTATCCCTTCTTGACCTATGAATTTCTAGTTGGTTCTCAGTAGCCCGGGGGAATA 420	
Db	7360	GGTTTATCCCTTCTTGACCTATGAATTTCTAGTTGGTTCTCAGTAGCCCGGGGGAATA 7419	
Qy	421	ATAGTAACAACAGCCATGATTTAGTGTAAATTTCTTGGTCTGGGCAGTGTCTCCTTTA 480	
Db	7420	ATAGTAACAACAGCCATGATTTAGTGTAAATTTCTTGGTCTGGGCAGTGTCTCCTTTA 7479	
Qy	481	ATCCTCAGAAACAACACTATGGATAGGTACAATTTATCTCTCACTTAAACAGATAAGAAACT 540	
Db	7480	ATCCTCAGAAACAACACTATGGATAGGTACAATTTATCTCTCACTTAAACAGATAAGAAACT 7539	
Qy	541	GAGCTCTCAGAAGGCTGAGCTATTTGCCCAAGATCACACAGCTTCTAAGTGGTGACAGTTT 600	
Db	7540	GAGCTCTCAGAAGGCTGAGCTATTTGCCCAAGATCACACAGCTTCTAAGTGGTGACAGTTT 7599	
Qy	601	GGGTTTTTTTTTTGTTGTTTAGAGACAGGGTCTTCTGTCAACCACGCATGAGCAC 660	
Db	7600	GGGTTTTTTTTTTGTTGTTTAGAGACAGGGTCTTCTGTCAACCACGCATGAGCAC 7659	
Qy	661	AGTGTGCAACCATAGGTCACCTGCAGCCTCAACCTCCTCTGAG 701	
Db	7660	AGTGTGCAACCATAGGTCACCTGCAGCCTCAACCTCCTCTGAG 7700	

RESULT 2

ABK47376	ABK47376 standard; DNA; 13612 BP.
ID	ABK47376;
AC	ABK47376;
XX	
DT	18-JUN-2002 (first entry)
XX	
DE	Human Phospholipase A2, groupIB (PLA2G1B) gene.
XX	
KW	Human; ds; gene; SNP; single nucleotide polymorphism; pancreatitis;
KW	pancreatic cancer; Phospholipase A2 groupIB; PLA2G1B; gene therapy;
KW	haplotype; genotype; chromosome 12q23-q24.1; transgenic; drug screening.
XX	
OS	Homo sapiens.
XX	
FH	Key
FT	variation
FT	Location/Qualifiers
FT	replace(3845,A)
FT	/*tag= a
FT	/label= SNP
FT	/note= "Single nucleotide polymorphism"
FT	variation
FT	replace(3968,A)
FT	/*tag= b
FT	/label= SNP
FT	/note= "Single nucleotide polymorphism"
FT	CDS
FT	4053..9613
FT	/*tag= c
FT	/product= "Pla2G1B"
FT	exon
FT	4053..4086
FT	/*tag= d
FT	/number= 1
FT	intron
FT	4087..5785

```
FT /tag= e
FT /number= 1
FT exon 5786..5945
FT /tag= f
FT /number= 2
FT intron 5946..6744
FT /tag= g
FT /number= 2
FT variation replace(6060,A)
FT /tag= h
FT /label= SNP
FT /note= "Single nucleotide polymorphism"
FT exon 6745..6873
FT /tag= i
FT /number= 3
FT variation replace(6844,A)
FT /tag= j
FT /label= SNP
FT intron /note= "Single nucleotide polymorphism"
FT 6874..9489
FT /tag= k
FT /number= 3
FT exon 9490..9613
FT /tag= l
FT /number= 4
FT variation replace(9531,A)
FT /tag= m
FT /label= SNP
FT /note= "Single nucleotide polymorphism"
```

WO200212562-A2.

14-FEB-2002.

06-AUG-2001; 2001WO-US024663.

04-AUG-2000; 2000US-0223179P.

(GENA-) GENA/ISSANCE PHARM INC.

Kazemi A, Kliem SE, Koshy B;

WPI; 2002-303982/34.

P-ESDB; AAU78667.

Novel isolated human Phospholipase A2, Group IB pancreas polynucleotide, for therapeutic purposes, for studying expression and function of the polynucleotide and for expressing the phospholipase protein.

Claim 1; Fig 1; 5lpp; English.

The invention relates to an isolated human Phospholipase A2, Group IB (pancreas) (PLA2G1B) polynucleotide comprising a sequence which is a polymorphic variant for a reference sequence for the PLA2G1B gene or its fragment, or a polymorphic variant of a reference sequence for a PLA2G1B cDNA or its fragment. Also included are haplotyping/genotyping the PLA2G1B gene of an individual, predicting the haplotype pair for the PLA2G1B gene of an individual, identifying an association between a trait and at least one haplotype or haplotype pair of the PLA2G1B gene, an isolated genotyping oligonucleotide for detecting a polymorphism in the PLA2G1B gene, a recombinant non-human organism transformed or transfected with the PLA2G1B sequence, where the organism expresses a PLA2G1B protein encoded by the first nucleotide sequence or by the polymorphic variant sequence, an isolated polypeptide comprising a sequence which is a polymorphic variant of a reference sequence for the PLA2G1B protein or its fragment, an anti-PLA2G1B monoclonal antibody, screening for drugs targeting PLA2G1B, a computer system for storing and analysing polymorphism data for the PLA2G1B gene and a genome anthology for PLA2G1B gene. The PLA2G1B variant is useful in studying the expression and function of PLA2G1B, and in expressing PLA2G1B protein for use in screening for candidate drugs to treat diseases related to PLA2G1B activity (e.g. pancreatitis and pancreatic cancer) and for therapeutic purposes. The transgenic organism is useful for studying expression of

CC the PLA2G1B isogenes in vivo, for in vivo screening and testing of drugs
CC targeted against PLA2G1B protein, and for testing the efficacy of
CC therapeutic agents and compounds in a biological system. The antibody is
CC useful for studying the effect of the variation on the biological
CC activity of PLA2G1B as well as on the binding affinity of candidate drugs
CC targeting PLA2G1B. The present sequence is the PLA2G1B gene which is
CC located on chromosome 12q23-q24.1
XX

SQ Sequence 13612 BP; 3637 A; 3290 C; 3070 G; 3615 T; 0 U; 0 Other;
Query Match 100.0%; Score 699.8; DB 6; Length 13612;
Best Local Similarity 99.6%; Pred. No. 4.7e-191;
Matches 698; Conservative 3; Mismatches 0; Indels 0; Gaps 0;

Qy	1	GTCTGTCTGCTGTCTCCAGCTGGGTAAACAGAGCAACTCTGTCTCAAAAAAATG	60
Db	6516	GTCTGTCTGCTGTCTCCAGCTGGGTAAACAGAGCAACTCTGTCTCAAAAAAATG	6575
Qy	61	CTTTCAATAAATATATGATAAAAGGACTTATATTTTCAAGCCATAGGATCATTTCTCC	120
Db	6576	CTTTCAATAAATATATGATAAAAGGACTTATATTTTCAAGCCATAGGATCATTTCTCC	6635
Qy	121	TGAAGCATCTTGGCGAAGTCATCCACCTGTTCCTGAGAGTGGGCGAGGTGAGGCTGAC	180
Db	6636	TGAAGCATCTTGGCGAAGTCATCCACCTGTTCCTGAGAGTGGGCGAGGTGAGGCTGAC	6695
Qy	181	CTATTGCTCTGCACTTACTCTATCTCAGCTGCTCCCTCCACCTTCCAGGTGCTGCAGA	240
Db	6696	CTATTGCTCTGCACTTACTCTATCTCAGCTGCTCCCTCCACCTTCCAGGTGCTGCAGA	6755
Qy	241	CACATGACAACTGCTAYGACAGGCCAAGAGCTGGACAGCTGTAATTTCTCTGGACA	300
Db	6756	CACATGACAACTGCTATGACAGGCCAAGAGCTGGACAGCTGTAATTTCTCTGGACA	6815
Qy	301	MMCCGTACACCCACACTATTTCATATCTGTGCTCTGGCTCGGCAATCACCTGTAGCAGTA	360
Db	6816	ACCCGTACACCCACACTATTTCATATCTGTGCTCTGGCTCGGCAATCACCTGTAGCAGTA	6875
Qy	361	GGTTATCCCTCCCTTGACCTATGAATTCCTAGTTGCTTCTCAGTAGGCCGGGGGAATA	420
Db	6876	GGTTATCCCTCCCTTGACCTATGAATTCCTAGTTGCTTCTCAGTAGGCCGGGGGAATA	6935
Qy	421	ATAGTAACAACAGCCCATGATTTAGTGTAAATTTCTTGGTTCTGGCAGTGTCTCCTTTA	480
Db	6936	ATAGTAACAACAGCCCATGATTTAGTGTAAATTTCTTGGTTCTGGCAGTGTCTCCTTTA	6995
Qy	481	ATCCTCAGAACACACTATGGGATAGGTACAATTATCTCACTTAAACAGATAAGAAAACT	540
Db	6996	ATCCTCAGAACACACTATGGGATAGGTACAATTATCTCACTTAAACAGATAAGAAAACT	7055
Qy	541	GAGGCTCAGAAGGCTGAGCTATTTGCCCAAGATCAACAGCTTGTAAAGTGGTCACAGTTT	600
Db	7056	GAGGCTCAGAAGGCTGAGCTATTTGCCCAAGATCAACAGCTTGTAAAGTGGTCACAGTTT	7115
Qy	601	GGGTTTTTTTTTTGTTGTTTGTAGACAGAGGCTTGTCTCTGTCAACCCAGGATGAGCAC	660
Db	7116	GGGTTTTTTTTTTGTTGTTTGTAGACAGAGGCTTGTCTCTGTCAACCCAGGATGAGCAC	7175
Qy	661	AGTGTGTCAACCATAGGTCACTGACGCTCAACCTCTCTGAG	701
Db	7176	AGTGTGTCAACCATAGGTCACTGACGCTCAACCTCTCTGAG	7216

RESULT 3

AD135082

ID AD135082 standard; DNA; 12174 BP.

XX

AC AD135082;

XX

DT 22-APR-2004 (first entry)

XX

DE Human PLA2G1B nucleotide sequence.

XX

CC antiasthmatic, cytostatic and analgesic activities. The compositions are
CC useful for the treatment of diseases associated with inflammation,
CC impaired airways, including lung disease and diseases whose secondary
CC effects afflict the lungs of a subject. They can be used for treating
CC e.g. ischaemic conditions, pulmonary vasoconstriction, allergies, asthma,
CC impeded respiration, respiratory distress syndrome, pain, cystic
CC fibrosis, pulmonary hypertension, emphysema, chronic obstructive
CC pulmonary disease (COPD), and cancers such as leukaemias, lymphomas,
CC carcinomas, and cancers which may metastasise to the lungs, including
CC breast and prostate cancer. The reduction of the adenosine content of the
CC ONS reduces side effects. The A-containing ONS break down with the
CC release of deoxyadenosine which activates adenosine receptors causing
CC bronchoconstriction and inflammation. AAA32313 to AAA35312 represent the
CC nucleotide sequences given in the sequence listing from the present
CC invention, which correspond to SEQ ID NO:1 to 2815, and then the last 185
CC sequences are also called SEQ ID NO:1 to 185, but the sequences differ
CC from the previously named sequences. SEQ ID NO:11 to 1680 (AAA32323 to
CC AAA3392) are specifically claimed ONS from the present invention. N.B.
CC Sequences given in the disclosure of the present invention do not match
CC up with their corresponding SEQ ID NO: sequences given in the sequence
CC listing

XX Sequence 3375 BP; 837 A; 866 C; 764 G; 908 T; 0 U; 0 Other;

Query Match 54.4%; Score 380.8; DB 3; Length 3375;

Best Local Similarity 99.2%; Pred. No. 4.1e-99;

Matches 379; Conservative 3; Mismatches 0; Indels 0; Gaps 0;

QY 1 GTCGTGTCACCTGCTGCAGCTGGGTAAACAGAGCACTCTGTCTCAAAAAAATG 60
DB ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
2994 GTCGTGTCACCTGCTGCAGCTGGGTAAACAGAGCACTCTGTCTCAAAAAAATG 3053
QY 61 CTTTCAATAATATATGATAAAGGACTTATATTTTCAAGCCATAGGATCATTTCTCC 120
DB ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
3054 CTTTCAATAATATATGATAAAGGACTTATATTTTCAAGCCATAGGATCATTTCTCC 3113
QY 121 TGAAGCATCTTGGCGAAGTCATCCCACTGTTCTGAGAGTGGCAGGTGAGGCTGAC 180
DB ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
3114 TGAAGCATCTTGGCGAAGTCATCCCACTGTTCTGAGAGTGGCAGGTGAGGCTGAC 3173
QY 181 CTATTGCTCGACTTACTCTTACTCTGCTGCTCCCTCCCACTTCCAGGTGCTGCCAGA 240
DB ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
3174 CTATTGCTCGACTTACTCTTACTCTGCTGCTCCCTCCCACTTCCAGGTGCTGCCAGA 3233
QY 241 CACATGACAACCTGTAYGACCAGGCCAAGAGCTGGACAGCTGTAAATTTCTGCTGGACA 300
DB ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
3234 CACATGACAACCTGTAGACACGACGACGACGACGACGACGACGACGACGACGAC 3293
QY 301 MWCCTACACCCACCACTTATCATCTGCTCTGCTCGGCTCGGCAATCACCTGTAGCAGTA 360
DB ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
3294 ACCGCTACACCCACCACTTATCATCTGCTCTGCTCGGCTCGGCAATCACCTGTAGCAGTA 3353
QY 361 GGTATTACCTTCTTGACCTA 382
DB ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
3354 GGTATTACCTTCTTGACCTA 3375

RESULT 5

AAF21223

ID AAF21223 standard; DNA; 3375 BP.

XX AAF21223;

AC AAF21223;

XX 14-MAR-2001 (first entry)

DE Human low adenosine antisense oligonucleotide related sequence #2790.

XX Low adenosine antisense oligonucleotide; phosphorothioate; allergy;
XX human; airway disorder; bronchoconstriction; lung inflammation;
KW surfactant depletion; respiratory; bronchodilator; antiinflammatory;
KW immunosuppressive; antiasthmatic; analgesic; hypotensive; cytostatic;
KW respiratory obstruction; pulmonary obstruction; impeded respiration;
KW surfactant hypoproduction; pulmonary vasoconstriction; asthma; RDS;

KW

KW

KW

KW

XX

OS

XX

PN

XX

XX

PD

XX

PF

XX

PR

XX

PA

XX

PA

XX

PI

XX

DR

XX

XX

PT

PT

XX

PS

XX

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

QY

Db

QY

Db

QY

121

TGAAGCATCTTGGCGAAGTCATCCCACTGTTCTGAGAGTGGCAGGTGAGGCTGAC 180

respiratory distress syndrome; pain; cystic fibrosis; allergic rhinitis;
pulmonary hypertension; emphysema; pulmonary transplantation rejection;
chronic obstructive pulmonary disease; pulmonary infection; bronchitis;
cancer; ss.

Homo sapiens.

WO200062736-A2.

26-OCT-2000.

24-MAR-2000; 2000WO-US008020.

06-APR-1999; 99US-0127958P.

(UYEC-) UNIV EAST CAROLINA.

(NYCE/) NYCE J W.

Nyce JW;

WPI; 2000-679539/66.

Low adenosine (A) content antisense oligonucleotides which do not trigger
adenosine receptors during metabolism, useful e.g. for treating cancers
and respiratory obstructions.
Disclosure; Page 1135-1136; 1592pp; English.

The present invention describes low adenosine (A) content antisense
oligonucleotides and compositions (i) comprising them. In the antisense
oligonucleotides the A is replaced by a 'Universal' or alternative base.
(i) can have respiratory, bronchodilator, antiinflammatory, analgesic,
immunosuppressive, antiasthmatic, hypotensive and cytostatic activities.
The antisense oligonucleotides and (i) can be used to down-regulate the
expression and or activity of target polypeptides associated with
lung/respiratory disorders and malignancies, such as stimulating and
activating peptide factors and transmitters, transcription factors,
immunoglobulins and antibodies, antibody receptors, cytokines and
chemokines, endogenously produced specific and non-specific enzymes,
binding proteins, adhesion molecules and their receptors, cytokine and
chemokine receptors, adenosine receptors, bradykinin receptors, central
nervous system (CNS) and peripheral nervous and non-nervous system
receptors, CNS and peripheral nervous and non-nervous system peptide
transmitters, defensins, growth factors, vasoactive peptides and
receptors, binding proteins and malignancy associated proteins. The
antisense oligonucleotides may be used in this way to treat disorders
including respiratory obstruction (especially pulmonary obstruction
and/or bronchoconstriction) and/or lung inflammation, allergy(ies) and/or
surfactant hypoproduction which are associated with a disease or
condition selected from pulmonary vasoconstriction, inflammation,
allergies, asthma, impeded respiration, respiratory distress syndrome
(RDS), pain, cystic fibrosis (CF), allergic rhinitis (AR), pulmonary
hypertension, emphysema, chronic obstructive pulmonary disease (COPD),
pulmonary transplantation rejection, pulmonary infections, bronchitis,
and/or cancer. AAF18434 to AAF21543 represent human polynucleotide
fragments and antisense oligonucleotides used in the exemplification of
the present invention

Sequence 3375 BP; 837 A; 866 C; 764 G; 908 T; 0 U; 0 Other;

Query Match 54.4%; Score 380.8; DB 3; Length 3375;

Best Local Similarity 99.2%; Pred. No. 4.1e-99;

Matches 379; Conservative 3; Mismatches 0; Indels 0; Gaps 0;

QY 1 GTCGTGTCACCTGCTGCAGCTGGGTAAACAGAGCACTCTGTCTCAAAAAAATG 60

Db ||||||||||||||||||||||||||||||||||||||||||||||||||||||||

2994 GTCGTGTCACCTGCTGCAGCTGGGTAAACAGAGCACTCTGTCTCAAAAAAATG 3053

QY 61 CTTTCAATAATATGATAAAGGACTTATATTTTCAAGCCATAGGATCATTTCTCC 120

Db ||||||||||||||||||||||||||||||||||||||||||||||||||||||||

3054 CTTTCAATAATATGATAAAGGACTTATATTTTCAAGCCATAGGATCATTTCTCC 3113

QY 121 TGAAGCATCTTGGCGAAGTCATCCCACTGTTCTGAGAGTGGCAGGTGAGGCTGAC 180

3114	TTGAAGCATTTGGGAAAGTATCCCACTGTGTCTTGAGTGGGAGGTGAGGGCTGAC	3173
181	CTATTGCTCTGCACCTTACTCTCTATCTCAGTGTCTCCCACTTTCCAGGTGCTGCAGA	240
3174	CTATTGCTCTGCACCTTACTCTCTATCTCAGTGTCTCCCACTTTCCAGGTGCTGCAGA	3233
241	CACATGACAACTGCTAYGACCAGGCCAAGAGCTGGACAGCTCTAAATTTCTCTCTGACA	300
3234	CACATGACAACTGCTATGACGACCGAGCCAAAGCTTGACAGCTGTAAATTTCTCTCTGACA	3293
301	MMCCGTACACCCACACACTATTCTATCTCGTGTCTCTGGCTCGGCAATCACCTGTAGCAGTA	360
3294	ACCCGTACACCCACACCTATTCTATCTCGTGTCTCTGGCTCGGCAATCACCTGTAGCAGTA	3353
361	GGTTTATCCCTTCTCTTGACCTA	382
3354	GGTTTATCCCTTCTCTTGACCTA	3375

RESULT 6	
ABZ96917	
ID	ABZ96917 standard; DNA; 3375 BP.
XX	
XX	
XX	ABZ96917;
XX	
XX	
DT	17-OCT-2003 (first entry)
XX	
XX	
DE	Human nucleic acid sequence.
XX	
XX	Human; antisense; lung dysfunction; nasal airway dysfunction;
KW	antiinflammatory steroid; ubiquinone; antiinflammatory; antiallergic;
KW	antisaethmatic; hypotensive; immunosuppressive; cytostatic; gene therapy;
KW	antisense gene therapy; respiratory; lung; adenosine sensitivity;
KW	adenosine receptor; bronchodilation; bronchoconstriction; lung allergy;
KW	lung inflammation; respiratory disease; ds.
XX	
OS	Homo sapiens.
XX	
PN	WO200285308-A2.
XX	
PD	31-OCT-2002.
XX	
XX	
PF	23-APR-2002; 2002WO-US013135.
XX	
PR	24-APR-2001; 2001US-0286137P.
XX	
PA	(EPIG-) EPIGENESIS PHARM INC.
XX	
PI	Nyce JW, Li Y, Sandraseagra A, Katz E, Pabalan J, Aguilar D;
PI	Miller S, Tang L, Shahabuddin S;
XX	
DR	WPI; 2003-229219/22.

Pharmaceutical composition for treating ailments associated with impaired respiration, has oligo(s) antisense to specific gene(s) or its corresponding RNAs, and glucocorticoid or non-glucocorticoid steroid or ubiquinone.

Disclosure; SEQ ID NO 12159; 872pp; English.

The invention relates to a novel pharmaceutical composition, which has a first active agent comprising an oligonucleotide antisense to the initiation codon, coding region, 5' or 3' and genomic flanking regions, 5' and 3' intron-exon junctions, or regions within 2-10 nucleotides of junctions of genes encoding a polypeptide associated with lung and/or nasal airway dysfunction and a second active agent comprising an antiinflammatory steroid and ubiquinone. A composition of the invention has antiinflammatory, antiallergic, antiasthmatic, hypotensive, have a immunosuppressive, and cytostatic activity. The composition may have a use in antisense gene therapy. The composition is useful for treating or preventing a respiratory, lung or malignant disease or condition, also for enhancing the prophylactic or therapeutic respiratory effect of an

[illegible]

RESULT 7	
ABD20766	
ID	ABD20766 standard; DNA; 3375 BP.
XX	
AC	ABD20766;
XX	
DT	29-JUL-2004 (first entry)
XX	
DE	Human pulmonary and inflammatory target DNA #377.
XX	
KW	Human; antisense; bronchoconstriction; allergy; hyposecretion; pain;
KW	respiratory tract inflammation; adenosine sensitivity; lung; cancer;
KW	surfactant depletion; anti-allergic; anti-inflammatory; antiasthmatic;
KW	analgesic; hypotensive; immunosuppressive; cytostatic; cystic fibrosis;
KW	beta-adrenergic agonist; respiratory disease; pulmonary vasoconstriction;
KW	respiratory distress syndrome; allergic rhinitis; pulmonary hypertension;
KW	emphysema; chronic obstructive pulmonary disease; cancer; bronchitis;
XX	pulmonary transplantation rejection; ds.
XX	
OS	Homo sapiens.
XX	
PN	WO200285309-A2.
XX	
PD	31-OCT-2002.
XX	
PF	23-APR-2002; 2002WO-US013143.
PR	24-APR-2001; 2001US-0286036P.
XX	
XX	

CC	listing	
XX	Sequence	6771 BP; 1815 A; 1570 C; 1439 G; 1947 T; 0 U; 0 Other;
SQ	Query Match	54.4%; Score 380.8; DB 3; Length 6771;
	Best Local Similarity	99.2%; Pred. No. 5.6e-99;
	Matches 379; Conservative	3; Mismatches 0; Indels 0; Gaps 0;
QY	1	GTCTGTCACCTGCTGCTCCAGCGCTGGTAAACAGAGCAACTCTGTCTCAAAAAAATG 60
Db	3556	GTCTGTCACCTGCTGCTCCAGCGCTGGTAAACAGAGCAACTCTGTCTCAAAAAAATG 3615
QY	61	CTTTCAATAATATATGATAAAGCACTATATTTTCAAGCCATAGGATCATTTCTCC 120
Db	3616	CTTTCAATAATATATGATAAAGCACTATATTTTCAAGCCATAGGATCATTTCTCC 3675
QY	121	TGAAGCATCTTGGCGAAGTCATCCCACTGTTCTGAGAGTGGCGAGGTGAGGGCTGAC 180
Db	3676	TGAAGCATCTTGGCGAAGTCATCCCACTGTTCTGAGAGTGGCGAGGTGAGGGCTGAC 3735
QY	181	CTATTGCTCTGCATTTACTCTCTATCTCAGCTGTCTCCCTCCCACTTTTCCAGGTGCTGCCAGA 240
Db	3736	CTATTGCTCTGCATTTACTCTCTATCTCAGCTGTCTCCCTCCCACTTTTCCAGGTGCTGCCAGA 3795
QY	241	CACATGACAACTGCTAYGACAGGCAAGCAAGCTGCAAGCTGCTAAATTTCTGCTGGACA 300
Db	3796	CACATGACAACTGCTAYGACAGGCAAGCAAGCTGCTAAATTTCTGCTGGACA 3855
QY	301	MMCCGTACACCCACCACTTATTCATCTCGTGTCTGGCTCGGCAATACCTCTAGCAGTA 360
Db	3856	ACCCGTACACCCACCACTTATTCATCTCGTGTCTGGCTCGGCAATACCTCTAGCAGTA 3915
QY	361	GGTTTATCCCTTCCTTGACCTA 382
Db	3916	GGTTTATCCCTTCCTTGACCTA 3937
RESULT 9		
ID	AAF21225	
XX	AAF21225 standard; DNA; 6771 BP.	
AC	AAF21225;	
XX		
DT	14-MAR-2001 (first entry)	
XX	Human low adenosine antisense oligonucleotide related sequence #2792.	
DE		
XX		
KW	Low adenosine antisense oligonucleotide; phosphorothioate; allergy;	
KW	human; airway disorder; bronchoconstriction; lung inflammation;	
KW	surfactant depletion; respiratory; bronchodilator; antiinflammatory;	
KW	immunosuppressive; antiasthmatic; analgesic; hypotensive; cytostatic;	
KW	respiratory obstruction; pulmonary vasoconstriction; impeded respiration;	
KW	surfactant hypoproduction; pulmonary vasoconstriction; asthma; RDS;	
KW	respiratory distress syndrome; pain; cystic fibrosis; allergic rhinitis;	
KW	pulmonary hypertension; emphysema; pulmonary transplantation rejection;	
KW	chronic obstructive pulmonary disease; pulmonary infection; bronchitis;	
XX	cancer; ss.	
OS	Homo sapiens.	
XX		
PN	WO20062736-A2.	
XX		
PD	26-OCT-2000.	
XX		
PF	24-MAR-2000; 2000WO-US008020.	
XX		
PR	06-APR-1999; 99US-0127958P.	
XX		
PA	(UYEC-) UNIV EAST CAROLINA.	
PA	(NYCE/) NYCE J W.	
XX		
PI	Nyce JW;	
XX		

DR	WPI; 2000-679539/66.	
XX		
PT	Low adenosine (A) content antisense oligonucleotides which do not trigger	
PT	adenosine receptors during metabolism, useful e.g. for treating cancers	
PT	and respiratory obstructions.	
XX		
PS	Disclosure; Page 1137-1139; 1592pp; English.	
XX		
CC	The present invention describes low adenosine (A) content antisense	
CC	oligonucleotides and compositions (I) comprising them. In the antisense	
CC	oligonucleotides the A is replaced by a 'Universal' or alternative base.	
CC	(I) can have respiratory, bronchodilator, antiinflammatory, analgesic,	
CC	immunosuppressive, antiasthmatic, hypotensive and cytostatic activities.	
CC	The antisense oligonucleotides and (I) can be used to down-regulate the	
CC	expression and or activity of target polypeptides associated with	
CC	lung/respiratory disorders and malignancies, such as stimulating and	
CC	activating peptide factors and transmitters, transcription factors and	
CC	immunoglobulins and antibodies, antibody receptors, cytokines and	
CC	chemokines, endogenously produced specific and non-specific enzymes,	
CC	binding proteins, adhesion molecules and their receptors, cytokine and	
CC	chemokine receptors, adenosine receptors, bradykinin receptors, central	
CC	nervous system (CNS) and peripheral nervous and non-nervous system	
CC	receptors, CNS and peripheral nervous and non-nervous system peptide	
CC	transmitters, defensins, growth factors, vasoactive peptides and	
CC	receptors, binding proteins and malignancy associated proteins. The	
CC	antisense oligonucleotides may be used in this way to treat disorders	
CC	including respiratory obstruction (especially pulmonary obstruction	
CC	and/or bronchoconstriction) and/or lung inflammation, allergy(ies) and/or	
CC	surfactant hypoproduction which are associated with a disease or	
CC	condition selected from pulmonary vasoconstriction, inflammation,	
CC	allergies, asthma, impeded respiration, respiratory distress syndrome	
CC	(RDS), pain, cystic fibrosis (CF), allergic rhinitis (AR), pulmonary	
CC	hypertension, emphysema, chronic obstructive pulmonary disease (COPD),	
CC	pulmonary transplantation rejection, pulmonary infections, bronchitis,	
CC	and/or cancer. AAF18434 to AAF21543 represent human polynucleotide	
CC	fragments and antisense oligonucleotides used in the exemplification of	
CC	the present invention	
XX		
SQ	Sequence 6771 BP; 1815 A; 1570 C; 1439 G; 1947 T; 0 U; 0 Other;	
	Query Match	54.4%; Score 380.8; DB 3; Length 6771;
	Best Local Similarity	99.2%; Pred. No. 5.6e-99;
	Matches 379; Conservative	3; Mismatches 0; Indels 0; Gaps 0;
QY	1	GTCTGTCACCTGCTGCTCCAGCGCTGGTAAACAGAGCAACTCTGTCTCAAAAAAATG 60
Db	3556	GTCTGTCACCTGCTGCTCCAGCGCTGGTAAACAGAGCAACTCTGTCTCAAAAAAATG 3615
QY	61	CTTTCAATAATATATGATAAAGCACTATATTTTCAAGCCATAGGATCATTTCTCC 120
Db	3616	CTTTCAATAATATATGATAAAGCACTATATTTTCAAGCCATAGGATCATTTCTCC 3675
QY	121	TGAAGCATCTTGGCGAAGTCATCCCACTGTTCTGAGAGTGGCGAGGTGAGGGCTGAC 180
Db	3676	TGAAGCATCTTGGCGAAGTCATCCCACTGTTCTGAGAGTGGCGAGGTGAGGGCTGAC 3735
QY	181	CTATTGCTCTGCATTTACTCTCTATCTCAGCTGTCTCCCTCCCACTTTTCCAGGTGCTGCCAGA 240
Db	3736	CTATTGCTCTGCATTTACTCTCTATCTCAGCTGTCTCCCTCCCACTTTTCCAGGTGCTGCCAGA 3795
QY	241	CACATGACAACTGCTAYGACAGGCAAGCAAGCTGCAAGCTGCTAAATTTCTGCTGGACA 300
Db	3796	CACATGACAACTGCTAYGACAGGCAAGCAAGCTGCTAAATTTCTGCTGGACA 3855
QY	301	MMCCGTACACCCACCACTTATTCATCTCGTGTCTGGCTCGGCAATACCTCTAGCAGTA 360
Db	3856	ACCCGTACACCCACCACTTATTCATCTCGTGTCTGGCTCGGCAATACCTCTAGCAGTA 3915
QY	361	GGTTTATCCCTTCCTTGACCTA 382
Db	3916	GGTTTATCCCTTCCTTGACCTA 3937

RESULT 10

ID ABZ96919
XX ABZ96919 standard; DNA; 6771 BP.
AC
XX ABZ96919;
DT 17-OCT-2003 (first entry)
XX
XX Human nucleic acid sequence.
DE
XX
XX Human; antisense; lung dysfunction; nasal airway dysfunction;
KW antiinflammatory steroid; ubiquinone; antiinflammatory; antiallergic;
KW antiasthmatic; hypotensive; immunosuppressive; cytostatic; gene therapy;
KW antisense gene therapy; respiratory; lung; adenosine sensitivity;
KW adenosine receptor; bronchodilation; bronchoconstriction; lung allergy;
KW lung inflammation; respiratory disease; ds.
XX
XX Homo sapiens.
XX
XX WO200285308-A2.
XX
XX 31-OCT-2002.
XX
XX 23-APR-2002; 2002WO-US013135.
XX
XX 24-APR-2001; 2001US-0286137P.
XX
XX (EPIG-) EPIGENESIS PHARM INC.
XX
XX Nyce JW, Li Y, Sandrasagra A, Katz E, Pabalan J, Aguilar D;
PI Miller S, Tang L, Shahabuddin S;
XX
XX WPI; 2003-229219/22.
XX
XX Pharmaceutical composition for treating ailments associated with impaired
PT respiration, has oligo(s) antisense to specific gene(s) or its
PT corresponding RNAs, and glucocorticoid or non-glucocorticoid steroid or
PT ubiquinone.
XX
XX Disclosure; SEQ ID NO 12161; 872pp; English.
XX
XX The invention relates to a novel pharmaceutical composition, which has a
CC first active agent comprising an oligonucleotide antisense to the
CC initiation codon, coding region, 5' or 3' end genomic flanking regions,
CC 5' and 3' intron-exon junctions, or regions within 2-10 nucleotides of
CC junctions of genes encoding a polypeptide associated with lung and/or
CC nasal airway dysfunction and a second active agent comprising an
CC antiinflammatory steroid and ubiquinone. A composition of the invention
CC has antiinflammatory, antiallergic, antiasthmatic, hypotensive,
CC immunosuppressive, and cytostatic activity. The composition may have a
CC use in antisense gene therapy. The composition is useful for treating or
CC preventing a respiratory, lung or malignant disease or condition, also
CC for enhancing the prophylactic or therapeutic respiratory effect of an
CC antiinflammatory steroid in a subject, for reducing or depleting levels
CC of, or reducing sensitivity to adenosine, reducing levels of adenosine
CC receptor, producing bronchodilation, increasing levels of ubiquinone or
CC lung surfactant in a subject's tissue, or treating bronchoconstriction,
CC lung inflammation, lung allergies, or a respiratory disease or condition.
CC Note: The sequence data for this patent is not represented in the printed
CC specification, but was obtained in electronic format directly from WIPO
CC at ftp.wipo.int/pub/published_pct_sequences
XX
XX Sequence 6771 BP; 1815 A; 1570 C; 1439 G; 1947 T; 0 U; 0 Other;
Query Match 54.4%; Score 380.8; DB 10; Length 6771;
Best Local Similarity 99.2%; Pred. No. 5.6e-99;
Matches 379; Conservative 3; Mismatches 0; Indels 0; Gaps 0;
OY 1 GTCGGTGCACCTGCTCCAGCTGGTAAACAGAGCACTCTCTCAAAAAAATG 60
DB 3556 GTCGGTGCACCTGCTCCAGCTGGTAAACAGAGCACTCTCTCAAAAAAATG 3615
OY 61 CTTTCAATAATATATGATAAAAGGACTTATATTTTCAAGCCATAGGATCATTTCTCC 120

DB 3616 CTTTCAATAATATATGATAAAAGGACTTATATTTTCAAGCCATAGGATCATTTCTCC 3675
OY 121 TGAAGCATCTTGGGAGAGTCAATCCCACTCTTCTGAGAGTGGCAGGTGAGGCTCAC 180
DB 3676 TGAAGCATCTTGGGAGAGTCAATCCCACTCTTCTGAGAGTGGCAGGTGAGGCTCAC 3735
OY 181 CTATTGCTCTGCACCTTACTCTCTATCTCAGCTGTCTCCCTCCACTTTTCCAGGTGCTCCAGA 240
DB 3736 CTATTGCTCTGCACCTTACTCTCTATCTCAGCTGTCTCCCTCCACTTTTCCAGGTGCTCCAGA 3795
OY 241 CACATGACAACTGTCTAGCAGCCCAAGAGCTGGACAGCTGTAAATTTCTGCTGGACA 300
DB 3796 CACATGACAACTGTCTAGCAGCCCAAGAGCTGGACAGCTGTAAATTTCTGCTGGACA 3855
OY 301 MMCCGTACACCCACCACTTATCTATCTCTGCTCTGGCTCGGCAATCACTGTAGCAGTA 360
DB 3856 ACCCGTACACCCACCACTTATCTATCTCTGCTCTGGCTCGGCAATCACTGTAGCAGTA 3915
OY 361 GGTATTATCCCTTCTCTTGACCTA 382
DB 3916 GGTATTATCCCTTCTCTTGACCTA 3937
RESULT 11
ABD20768
ID ABD20768 standard; DNA; 6771 BP.
XX
AC ABD20768;
XX
DT 29-JUL-2004 (first entry)
XX
XX Human pulmonary and inflammatory target DNA #379.
XX
XX Human; antisense; bronchoconstriction; allergy; hyposecretion; pain;
KW respiratory tract inflammation; adenosine sensitivity; lung; cancer;
KW surfactant depletion; antiallergic; antiinflammatory; antiasthmatic;
KW analgesic; hypotensive; immunosuppressive; cytostatic; cystic fibrosis;
KW beta-adrenergic agonist; respiratory disease; pulmonary vasoconstriction;
KW respiratory distress syndrome; allergic rhinitis; pulmonary hyperextension;
KW emphysema; chronic obstructive pulmonary disease; cancer; bronchitis;
KW pulmonary transplantation rejection; ds.
XX
XX Homo sapiens.
XX
XX WO200285309-A2.
XX
XX 31-OCT-2002.
XX
XX 23-APR-2002; 2002WO-US013143.
XX
XX 24-APR-2001; 2001US-0286036P.
XX
XX (EPIG-) EPIGENESIS PHARM INC.
XX
XX Nyce JW, Li Y, Sandrasagra A, Katz E, Pabalan J, Aguilar D;
PI Miller S, Tang L, Shahabuddin S;
XX
XX WPI; 2003-093058/08.
XX
XX Pharmaceutical composition for treating asthma, has antisense
PT oligonucleotide containing less percentage of adenosine, targeted to
PT nucleic acids associated with lung airway or lung dysfunction, and
PT bronchodilating agent.
XX
XX Claim 15; SEQ ID NO 12161; 763pp; English.
XX
XX This invention describes a novel composition (a) a first active agent,
CC comprising oligonucleotides, effective for alleviating
CC bronchoconstriction, respiratory tract inflammation, allergies and
CC reducing adenosine sensitivity, levels of adenosine (A) or (A) receptors,
CC surfactant depletion or hyposecretion, when administered to a mammal. The
CC oligonucleotides are derived from a gene encoding or regulating

QY 341 GGCAATCACCCTGTAGCAGTA 360
DB 305 GGCAATCACCCTGTAGCAGCA 324

RESULT 13
ID ABK47377 standard; cDNA; 447 BP.
AC
AC ABK47377;
DT 18-JUN-2002 (first entry)
XX
DE Human Phospholipase A2, group IB (PLA2G1B) cDNA.
KW Human; ss; gene; SNP; single nucleotide polymorphism; pancreatitis;
KW pancreatic cancer; Phospholipase A2 group IB; PLA2G1B; gene therapy;
KW haplotype; genotype; chromosome 12q23-q24.1; transgenic; drug screening.
OS Homo sapiens.
XX
XX Key Location/Qualifiers
FH 1. 447
FT CDS /tag= a
FT /product= "Pla2G1B"
FT replace(294,A)
FT variation /tag= h
FT /label= SNP
FT /note= "Single nucleotide polymorphism"
FT replace(365,A)
FT /tag= j
FT /label= SNP
FT /note= "Single nucleotide polymorphism"
XX
XX WO200212562-A2.
XX
XX 14-FEB-2002.
XX
XX 06-AUG-2001; 2001WO-US024663.
XX
XX 04-AUG-2000; 2000US-0223179P.
XX
XX (GNA-) GENAISSANCE PHARM INC.
XX
XX Kazemi A, Kliem SE, Koshy B;
XX
XX WPI; 2002-303982/34.
XX P-FSDB; AAU78667.
XX
XX Novel isolated human Phospholipase A2, Group IB pancreas polynucleotide,
XX for therapeutic purposes, for studying expression and function of the
XX polynucleotide and for expressing the phospholipase protein.
XX
XX Claim 27; Fig 2; 51pp; English.
XX
XX The invention relates to an isolated human Phospholipase A2, Group IB
XX (pancreas) (PLA2G1B) polynucleotide comprising a sequence which is a
XX polymorphic variant for a reference sequence for the PLA2G1B gene or its
XX fragment, or a polymorphic variant of a reference sequence for a PLA2G1B
XX cDNA or its fragment. Also included are haplotyping/genotyping the
XX PLA2G1B gene of an individual, predicting the haplotype pair for the
XX PLA2G1B gene of an individual, identifying an association between a trait
XX and at least one haplotype or haplotype pair of the PLA2G1B gene, an
XX isolated genotyping oligonucleotide for detecting a polymorphism in the
XX PLA2G1B gene, a recombinant non-human organism transformed or transfected
XX with the PLA2G1B sequence, where the organism expresses a PLA2G1B protein.
XX encoded by the first nucleotide sequence or by the polymorphic variant
XX sequence, an isolated polypeptide comprising a sequence which is a
XX polymorphic variant of a reference sequence for the PLA2G1B protein or
XX its fragment, an anti-PLA2G1B monoclonal antibody, screening for drugs
XX targeting PLA2G1B, a computer system for storing and analysing
XX polymorphism data for the PLA2G1B gene and a genome anthology for PLA2G1B
XX gene. The PLA2G1B variant is useful in studying the expression and

CC function of PLA2G1B, and in expressing PLA2G1B protein for use in
CC screening for candidate drugs to treat diseases related to PLA2G1B
CC activity (e.g. pancreatitis and pancreatic cancer) and for therapeutic
CC purposes. The transgenic organism is useful for studying expression of
CC the PLA2G1B isogenes in vivo, for in vivo screening and testing of drugs
CC targeted against PLA2G1B protein, and for testing the efficacy of
CC therapeutic agents and compounds in a biological system. The antibody is
CC useful for studying the effect of the variation on the biological
CC activity of PLA2G1B as well as on the binding affinity of candidate drugs
CC targeting PLA2G1B. The PLA2G1B gene is located on chromosome 12q23-q24.1.
CC The present sequence is the cDNA encoding PLA2G1B
XX
SQ Sequence 447 BP; 116 A; 128 C; 112 G; 91 T; 0 U; 0 Other;
Query Match 18.7%; Score 130.8; DB 6; Length 447;
Best Local Similarity 94.3%; Pred. No. 3e-27; Indels 0; Gaps 0;
Matches 132; Conservative 3; Mismatches 5;
QY 221 ACTTTCCAGGTGTCGCCAGACACATGACAACCTGCTAYGACCGCCAGAGCTGGACAG 280
DB 186 ACTGGACAAGTGTCTGCCAGACACATGACAACCTGCTATGACCGCCAGAGCTGGACAG 245
QY 281 CTGTAAATTTCTGCTGGACAMCCGTACACCCACACCTATTTCATCTGCTCTGGCTC 340
DB 246 CTGTAAATTTCTGCTGGACACACCGGTACACCCACCTATTTCATCTGCTCTGGCTC 305
QY 341 GGCAATCACCCTGTAGCAGTA 360
DB 306 GGCAATCACCCTGTAGCAGCA 325

RESULT 14
ADV43454
ID ADV43454 standard; cDNA; 447 BP.
XX
XX AC ADV43454;
XX
XX 10-MAR-2005 (first entry)
XX
XX Human psychoneuroendocrine-immune expressed sequence tag SEQ ID NO 1082.
DE
XX microarray; psychoneuroendocrine-immune; chronic fatigue;
KW non-insulin dependent diabetes; allergy; immune disorder; inflammation;
KW cancer; neoplasm; infection; expressed sequence tag; ss.
XX
XX Homo sapiens.
XX
XX WO2004108899-A2.
XX
XX 16-DEC-2004.
XX
XX 04-JUN-2004; 2004WO-US017686.
XX
XX 04-JUN-2003; 2003US-0475915P.
XX
XX (USSH) US DEPT HEALTH & HUMAN SERVICES.
XX
XX Nicholson A, Vernon SD;
XX
XX WPI; 2005-031682/03.
XX
XX New microarray comprising probes for genes involved in
XX psychoneuroendocrine-immune (PNI) activity, useful in diagnosing a
XX condition associated with PNI activity, e.g., inflammatory or infectious
XX diseases.
XX
XX Claim 1; SEQ ID NO 1082; 254pp; English.
XX
XX The invention relates to a new microarray which comprises probes for
XX genes involved in psychoneuroendocrine-immune (PNI) activity. The
XX microarray is useful in diagnosing a condition associated with PNI
XX activity, such as CFS, type-2 diabetes, allergic condition, inflammation,
XX cancer and infection. The present sequence represents a

CC psychoneuroendocrinimmune gene expressed sequence tag. Note the
 CC specification mentions SEQ ID NO of up to 314 but only sequences up to
 CC SEQ ID NO 1829 are provided.
 XX
 SQ Sequence 447 BP; 116 A; 128 C; 112 G; 91 T; 0 U; 0 Other;
 Query Match 18.7%; Score 130.8; DB 14; Length 447;
 Best Local Similarity 94.3%; Pred. No. 3e-27; Mismatches 3; Indels 0; Gaps 0;
 Matches 132; Conservative 3;
 QY 221 ACTTTCCAGGTGCTGCCAGACACATGACAACCTGCTAYGACCAAGGCGCAAGAGCTGGACAG 280
 DB 186 ACTGGACAAGTGTCTGCCAGACACATGACAACCTGCTATGACCAAGGCGCAAGAGCTGGACAG 245
 QY 281 CTGTAAATTTCTGTGGACAMMCGTACACCCACCTATTTCATCTCTGCTCTGGCTC 340
 DB 246 CTGTAAATTTCTGTGGACAMMCGTACACCCACCTATTTCATCTCTGCTCTGGCTC 305
 QY 341 GGCAATCACCTGTAGCAGTA 360
 DB 306 GGCAATCACCTGTAGCAGCA 325

RESULT 15

ID AAA35100 standard; DNA; 562 BP.
 AC AAA35100;

XX
 DT 28-JUL-2000 (first entry)
 XX
 DE Human adenosine receptor related polynucleotide SEQ ID NO:2789.
 XX
 KW Human; adenosine receptor; low adenosine antisense oligonucleotide;
 KW phosphorothioate; impaired respiration; inflammation; allergy;
 KW allergic disease; bronchoconstriction; inhibitor; antiinflammatory;
 KW antiallergic; antiaslathmatic; cytostatic; analgesic; impaired airway;
 KW lung disease; ischaemic condition; pulmonary vasoconstriction; asthma;
 KW respiratory distress syndrome; pain; cystic fibrosis; emphysema;
 KW pulmonary hypertension; chronic obstructive pulmonary disease; COPD;
 KW cancer; leukaemia; lymphoma; carcinoma; metastasis; ss.
 XX
 OS Homo sapiens.
 XX
 XX WO200009525-A2.
 XX
 XX 24-FEB-2000.
 XX
 XX 03-AUG-1999; 99WO-US017712.
 XX
 XX 03-AUG-1998; 98US-0095212P.
 XX
 XX (UYEC-) UNIV EAST CAROLINA.
 XX
 XX Nyce JW;
 XX
 XX WPI; 2000-205971/18.
 XX
 XX New antisense oligonucleotides useful for treating e.g. pulmonary
 PT vasoconstriction, inflammation, allergies, asthma, hypertension,
 PT bronchitis, emphysema, respiratory distress syndrome, ischemia or
 PT cancers.
 XX
 XX Disclosure; Page 1056; 1343pp; English.
 XX
 CC The present invention describes a new composition comprising an antisense
 CC oligonucleotide (ON) with low adenosine (up to 15%), which targets
 CC nucleic acids involved in bronchoconstriction, allergies, and/or
 CC inflammation. The ON can have antiinflammatory, antiallergic,
 CC antiaslathmatic, cytostatic and analgesic activities. The compositions are
 CC useful for the treatment of diseases associated with inflammation,
 CC impaired airways, including lung disease and diseases whose secondary
 CC effects afflict the lungs of a subject. They can be used for treating

CC e.g. ischaemic conditions, pulmonary vasoconstriction, allergies, asthma,
 CC impeded respiration, respiratory distress syndrome, pain, cystic
 CC fibrosis, pulmonary hypertension, emphysema, chronic obstructive
 CC pulmonary disease (COPD), and cancers such as leukaemias, lymphomas,
 CC carcinomas, and cancers which may metastasise to the lungs, including
 CC breast and prostate cancer. The reduction of the adenosine content of the
 CC ONs reduces side effects. The A-containing ONs break down with the
 CC release of deoxyadenosine which activates adenosine receptors causing the
 CC bronchoconstriction and inflammation. AAA32313 to AAA35312 represent the
 CC nucleotide sequences given in the sequence listing from the present
 CC invention, which correspond to SEQ ID NO:1 to 2815, and then the last 185
 CC sequences are also called SEQ ID NO:1 to 185, but the sequences differ
 CC from the previously named sequences. SEQ ID NO:11 to 1680 (AAA32323 to
 CC AAA33922) are specifically claimed ONs from the present invention. N.B.
 CC Sequences given in the disclosure of the present invention do not match
 CC up with their corresponding SEQ ID NO: sequences given in the sequence
 CC listing

SQ Sequence 562 BP; 147 A; 163 C; 125 G; 127 T; 0 U; 0 Other;

Query Match 18.7%; Score 130.8; DB 3; Length 562;

Best Local Similarity 94.3%; Pred. No. 3.3e-27;

Matches 132; Conservative 3; Mismatches 5; Indels 0; Gaps 0;

QY 221 ACTTTCCAGGTGCTGCCAGACACATGACAACCTGCTAYGACCAAGGCGCAAGAGCTGGACAG 280
 DB 223 ACTGGACAAGTGTCTGCCAGACACATGACAACCTGCTATGACCAAGGCGCAAGAGCTGGACAG 282
 QY 281 CTGTAAATTTCTGTGGACAMMCGTACACCCACCTATTTCATCTCTGCTCTGGCTC 340
 DB 283 CTGTAAATTTCTGTGGACAMMCGTACACCCACCTATTTCATCTCTGCTCTGGCTC 342

QY 341 GGCAATCACCTGTAGCAGTA 360

DB 343 GGCAATCACCTGTAGCAGCA 362

Search completed: February 9, 2006, 18:54:17
 Job time : 454.558 secs

Result No.	Query	Score	Query			DB	ID	Description
			Match	Length	No.			
C 1	181.2	25.9	729	10	CE721339	CE721339	tigr-g8s-	
C 2	156.8	22.4	653	10	CE574035	CE574035	tigr-g8s-	
C 3	130.8	18.7	425	1	AM136966	AM136966	UI-H-Bil-	
C 4	130.8	18.7	459	1	AA845006	AA845006	ak61f09.s	
C 5	130.8	18.7	495	2	BG142111	BG142111	ia49b01.y	
C 6	130.8	18.7	497	1	A1985509	A1985509	ws08d11.x	
C 7	130.8	18.7	520	5	BU952425	BU952425	io74d05.x	
C 8	130.8	18.7	522	1	AW583679	AW583679	ia03a06.y	
C 9	130.8	18.7	526	3	BQ271518	BQ271518	ik15a01.x	
C 10	130.8	18.7	529	1	AA835293	AA835293	ak66b09.s	
C 11	130.8	18.7	529	3	B0101169	B0101169	ij24h05.y	
C 12	130.8	18.7	532	1	AA835315	AA835315	ak66e08.s	
C 13	130.8	18.7	537	5	BU950978	BU950978	io74d05.y	
C 14	130.8	18.7	539	1	AA844927	AA844927	ak61a06.s	
C 15	130.8	18.7	552	3	BM264109	BM264109	ig31a01.y	
C 16	130.8	18.7	562	1	A1925956	A1925956	wh12g03.x	
C 17	130.8	18.7	562	3	BP231552	BP231552	BP323152	
C 18	130.8	18.7	577	4	BC013384	BC013384	Homo sapi	
C 19	130.8	18.7	579	3	BP322008	BP322008	BP322008	
C 20	130.8	18.7	580	3	BP323927	BP323927	BP323927	
C 21	130.8	18.7	581	3	BP324031	BP324031	BP324031	
C 22	130.8	18.7	581	3	BP324615	BP324615	BP324615	

Db	573	GAATAATCTTTTAAAAATACATAAATAATAATAATAAATAAACCTGATCAAGGATTTAAGTT	514
Qy	94	TTTTTCAAGCATAAGGATC-ATTTCTCTCTGAAGCATCTTGGCGAAGTCATCCCCACCTGT	152
Db	513	TTTTTTGAGCCACAGGACCAATACTCTGGAGCATGTCCACTAAATCATCCCTCCCTAC	454
Qy	153	TCCTGAGAGTGGGACAGGTGAGGGCTGACCTATTGCTGTGACACTTACTCTCTATCTCAGCTG	212
Db	453	CTATTCTCTG-----AGGCTGAGCTGCTTCCCTACCCCTTC-----CTCAGCCCTG	409
Qy	213	TCCCTCCACATTTTCAGGTGTCGACACATGACAACTGCTAYGACCGCCGCAAGAG	272
Db	408	TTCTTCTCCCTCTCCAGGTGTTGCGAGCGCATGACCACTGCTACTCAGAAGCCAGAAA	349
Qy	273	CTGACAGCTGTAAATTTCTGCTGACACMMCGGTACACCCACCACTATTCTATCTCGTGC	332
Db	348	CTGGACAGCTGTAAATTTCTCTGACAAACCCCTACACAAAATCTACTCATCTCATGC	289
Qy	333	TCTGCTCGGCAATCACTGTAGCAGTAGGTTTATCCCTTCCCTTGACACCTATGAATTTCTAG	392
Db	288	TCCGCTCTGAGATCACCTGCAGCAGTAGGTTTATCCCTTGGCTTGTGAGCCCTCA	229
Qy	393	TTGGTTCTCAGTAGCGCGGGGGAATAATAGTAAACAGCCATGATTTAGTGTTAATT	452
Db	228	GTAGACAT-----GGGAGGAAGAATAATAATGACATCCACATGATGATCAATCACA	178
Qy	453	TTCTTTGTTCTGGCAGTGTCTCTTTAATCTCAGAAACACATATGGGATAGGTACAA	512
Db	177	TCCTTGGTGTAGGCTGTGTCCTCTTGAATCTTGGCAACAGCCCTAGGTGATCGGAATTA	118
Qy	513	TTATCTCT-CACTTAACAGATAAGAAAACCTGAGGCTCAGAAGCTGAGCTATTTGCCCAAG	571
Db	117	TTATTATATCCATTTTACAGATGACAAAACAGCTCAGAAGGTTGAGCCAATTGCGATGG	58
Qy	572	ATCACACAGCTGTAAAGTGTGACAGTTTGGGTT	605
Db	57	GGCACACAGCTCGTAAATGGTGTGTTGGTTTT	24
RESULT 2			
CE574035/c			
LOCUS			
DEFINITION			
tigr-gss-dog-1700031102249 Dog Library Canis familiaris genomic,			
genomic survey sequence.			
ACCESSION			
VERSION			
KEYWORDS			
SOURCE			
ORGANISM			
REFERENCE			
AUTHORS			
TITLE			
JOURNAL			
PUBMED			
COMMENT			
The dog genome: survey sequencing and comparative analysis			
Science 301 (5641), 1898-1903 (2003)			
14512627			
Contact: Kirkness EF			
The Institute for Genomic Research			
Department of Eukaryotic Genomics, TIGR, 9712 Medical Center Drive,			
Rockville, MD 20850, USA			
Tel: 301-838-0200			
Fax: 301-838-0208			
Email: ekirknes@tigr.org			
Class: shotgun.			
Location/Qualifiers			
1. .653			
/organism="Canis familiaris"			
/mol_type="genomic DNA"			
/strain="Standard Poodle"			
/db_xref="taxon:9615"			
FEATURES			
source			
ORIGIN			
Query Match 22.4%; Score 156.8; DB 10; Length 653;			
Best Local Similarity 62.2%; Pred. No. 2.2e-25;			
Matches 336; Conservative 3; Mismatches 175; Indels 26; Gaps 5;			
/clone lib="Dog Library"			
/notes="Site 1: BstXI; Libraries were prepared from			
peripheral blood"			
Qy	34	GCAACTCTGTCTCAAAAAAATGCTTTCAATAAATATATATGATAAAGGACTTATAT	93
Db	540	GAATAATCTTTTAAAAAATACAAAATATAAATAAATAAACCCTGATCAAGGATTTAAGTT	481
Qy	94	TTTTTCAAGCATAAGGATC-ATTTCTCTCTGAAGCATCTTGGCGAAGTCATCCCCACCTGT	152
Db	480	TTTTTTGAGCCACAGGACCAATACTCTGGAGCATGTCCACTAAATCATCCCTCCCTAC	421
Qy	153	TCCTGAGAGTGGGAGGTGAGGCTGACCTATTGCTCTGCACTTACTCTATCTCAGCTG	212
Db	420	CTATTCTCTG-----AGGCTGAGCTGCTTCCCTACCCCTTC-----CTCAGCCCTG	376
Qy	213	TCCCTCCACATTTTCAGGTGTCGACACATGACAACTGCTAYGACCGCCGCAAGAG	272
Db	375	TTCTTCTCCCTCTCCAGGTGTGCCAGACGATGACCACTGCTACTCAGAAGCCAGAAA	316
Qy	273	CTGACAGCTGTAAATTTCTGCTGACAMMCGGTACACCCACCACTATTCTATCTCGTGC	332
Db	315	CTGGACAGCTGTAAATTTCTCTGACAAACCCCTACACAAAATCTACTCATCTCATGC	256
Qy	333	TCTGCTCGGCAATCACTGTAGCAGTAGGTTTATCCCTTCCCTTGACACCTATGAATTTCTAG	392
Db	255	TCCGCTCTGAGATCACCTGCAGCAGTAGGTTTATCCCTTGGCTTGTGAGCCCTCA	196
Qy	393	TTGGTTCTCAGTAGCGCGGGGGAATAAATAGTAAACAGCCATGATTTA-GTGTAAAT	451
Db	195	GTAGACAT-----GGGAGGAAGAATAAATGACATCCACCAATGATCAATCACA	145
Qy	452	TTTCTTGGTCTGGCAGTGTCTCTTTAATCTCAGAAACACACTATGGGATAGGTACA	511
Db	144	TCCCTTGGTGTAGGCTGTGCTCTCTTGAATCTTTGGCAACAGCCCTAGGTGATCGGAATT	85
Qy	512	ATTATCTCTCACTTAACAGATAAGAAAACCTGAGGCTCAGAAGGCTGAGCTATTTGCCCAAG	571
Db	84	ATTATTATACCATTTTACAGATGACAAAACAGCTCAGAAGGTTGAGCCAATTGACCATG	25
RESULT 3			
AW139696/c			
LOCUS			
DEFINITION			
UI-H-B11-adr-h-03-0-UI.s1 NCI_CGAP_Sub3 Homo sapiens cDNA clone			
IMAGE:2717813 3', mRNA sequence.			
ACCESSION			
VERSION			
KEYWORDS			
SOURCE			
ORGANISM			
REFERENCE			
AUTHORS			
TITLE			
JOURNAL			
COMMENT			
1 (bases 1 to 425)			
NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.			
National Cancer Institute, Cancer Genome Anatomy Project (CGAP),			
Tumor Gene Index			
Unpublished (1997)			
Contact: Robert Strausberg, Ph.D.			
Email: cgapbs-r@mail.nih.gov			
The sequence contained an oligo-dT track that was present in the			
oligonucleotide that was used to prime the synthesis of first			
strand cDNA and therefore this may represent a bonafide poly A			
tail. cDNA Library Preparation: M.B. Soares Lab Clone distribution:			
NCI-CGAP clone distribution information can be found through the			
I.M.A.G.E. Consortium/LLNL at:			
www.bio.llnl.gov/bbrp/image/image.html			

REFERENCE
AUTHORS
TITLE
JOURNAL
COMMENT

Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini;
Hominidae; Homo.
1 (bases 1 to 495)
Melton, D., Meadows, A., Clifton, S., Hillier, L., Marra, M., Pape, D.,
Wylie, T., Martin, J., Blistain, A., Schmitt, A., Theising, B.,
Ritter, E., Ronko, I., Bennett, J., Cardenas, M., Gibbons, M.,
McAnn, R., Cole, R., Tsgareishvili, R., Williams, T., Jackson, Y. and
Bowers, Y.
WashU-Harvard Pancreas EST Project
Unpublished (2000)
Contact: Douglas Melton, Klaus H. Kaestner, & Hiroshi Inoue
Endocrine Pancreas Consortium
Harvard University, Howard Hughes Medical Institute
Dept of Molecular and Cellular Biology, 7 Divinity Ave, Cambridge,
MA 02138
Tel: 617-495-1812
Fax: 617-495-8557
Email: dmelton@biohp.harvard.edu
Library was constructed by Dr. Douglas Melton DNA sequencing by:
Washington University Genome Sequencing Center For information on
obtaining a clone please contact: Juliana Brown
(brownjas.harvard.edu) This sequence now available from the IMAGE
consortium, for clone orders contact: info@image.llnl.gov
Seq primer: -40RP from Gibco
High quality sequence stop: 458.

FEATURES
source
1. .495
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/clone="IMAGE:5641656"
/sex="Both"
/tissue_type="Islets of Langerhans"
/dev_stage="Adult"
/lab_host="TOP10"
/clone_lib="Melton Human Islets HI21"
/note="Organ: Pancreas; Vector: pZero-2; Site 1: Not I;
Site 2: Xho I; Library constructed using SuperScript
Plasmid library kit (Life Technologies). cDNA made by
oligo-dT priming. Xho I site destroyed during cloning.
Size-selected by column fractionation; average insert
size 1.59 kb. Primary library, unamplified."

ORIGIN
Query Match 18.7%; Score 130.8; DB 2; Length 495;
Best Local Similarity 94.3%; Pred. No. 2.1e-19;
Matches 132; Conservative 3; Mismatches 5; Indels 0; Gaps 0;
QY 221 ACTTCCAGGTGTCGCAGACATGACAACTGCTAYGACCCAGGCCAAGCTGGACAG 280
DB 178 ACTGGACAAGTGTGTCGACACATGACAACTGCTATGACCCAGGCCAAGCTGGACAG 237
QY 281 CTGTAATTTCTGCTGGACAMCCGTACACCCACCTTATCATCTCTGCTCTGGCTC 340
DB 238 CTGTAATTTCTGCTGGACAMCCGTACACCCACCTTATCATCTCTGCTCTGGCTC 297
QY 341 GGCAATCACCTGTAGCAGTA 360
DB 298 GGCAATCACCTGTAGCAGCA 317

RESULT 6
AI985509/c
LOCUS
DEFINITION
AI985509
497 bp mRNA linear EST 31-AUG-1999
ws08dl1.x1 NCI CGAP Kid11 Homo sapiens cDNA clone IMAGE:2496597 3',
similar to gb:M21054 PHOSPHOLIPASE A2 PRECURSOR (HUMAN);, mRNA
sequence.
AI985509
GI:5812786
EST.
KEYWORDS
Homo sapiens (human)
Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini;
Hominidae; Homo.

REFERENCE
AUTHORS
TITLE
JOURNAL
COMMENT

Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini;
Hominidae; Homo.
1 (bases 1 to 497)
NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.
National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
Tumor Gene Index
Unpublished (1997)
Contact: Robert Strausberg, Ph.D.
Email: cgapbs-rc@mail.nih.gov
Tissue Procurement: Christopher Moskaluk, M.D., Ph.D., Michael R.
Emmert-Buck, M.D., Ph.D.
cDNA Library Preparation: M. Bento Soares, Ph.D.
cDNA Library Arrayed by: Greg Lennon, Ph.D.
DNA Sequencing by: Washington University Genome Sequencing Center
Clone distribution: NCI-CGAP clone distribution information can be
found through the I.M.A.G.E. Consortium/LLNL at:
www-bio.llnl.gov/bbrp/image/image.html
Seq primer: -40UP from Gibco
High quality sequence stop: 386.

FEATURES
source
1. .497
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/clone="IMAGE:2496597"
/lab_host="DH10B"
/clone_lib="NCI CGAP Kid11"
/note="Organ: kidney; Vector: pT7T3D-Pac (Pharmacia) with
a modified polylinker; Site 1: Not I; Site 2: Eco RI;
Plasmid DNA from the normalized library NCI CGAP Kid3 was
prepared, and ss circles were made in vitro. Following HAP
purification, this DNA was used as tracer in a subtractive
hybridization reaction. The driver was PCR-amplified cDNAs
from a pool of 5,000 clones made from the same library
(cloneIDs 1322376-1323911, 1456007-1456775, and
1500552-1502855). Subtraction by Bento Soares and M.
Fatima Bonaldo."

ORIGIN
Query Match 18.7%; Score 130.8; DB 1; Length 497;
Best Local Similarity 94.3%; Pred. No. 2.1e-19;
Matches 132; Conservative 3; Mismatches 5; Indels 0; Gaps 0;
QY 221 ACTTCCAGGTGTCGCAGACATGACAACTGCTAYGACCCAGGCCAAGCTGGACAG 280
DB 343 ACTGGACAAGTGTGTCGACACATGACAACTGCTATGACCCAGGCCAAGCTGGACAG 284
QY 281 CTGTAATTTCTGCTGGACAMCCGTACACCCACCTTATCATCTCTGCTCTGGCTC 340
DB 283 CTGTAATTTCTGCTGGACAMCCGTACACCCACCTTATCATCTCTGCTCTGGCTC 224
QY 341 GGCAATCACCTGTAGCAGTA 360
DB 223 GGCAATCACCTGTAGCAGCA 204

RESULT 7
BU952425/c
LOCUS
DEFINITION
BU952425
520 bp mRNA linear EST 21-OCT-2002
io74d05.x1 HR85 islet Homo sapiens cDNA clone IMAGE:6132129 3',
similar to SW:PA21_HUMAN P04054 PHOSPHOLIPASE A2 PRECURSOR ;, mRNA
sequence.
BU952425
GI:24204177
EST.
KEYWORDS
Homo sapiens (human)
Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini;
Hominidae; Homo.
1 (bases 1 to 520)
REFERENCE
AUTHORS
Lemishka, I., Searce, M., Brestelli, J., Gradwohl, G., Clifton, S.,
Melton, D., Brown, J., Kenty, G., Permutt, A., Lee, C., Kaestner, K.,
Permutt, A., Lee, C., Kaestner, K.,

Hillier, L., Marra, M., Pape, D., Wyllie, T., Martin, J., Blistain, A., Schmitt, A., Theising, B., Ritter, E., Ronko, I., Bennett, J., Cardenas, M., Gibbons, M., McCann, R., Cole, R., Tsagarishvili, R., Williams, T., Jackson, Y., and Bowers, Y.
 Unpublished (2000)
 Other ESTs: i07405.y1
 Contact: Douglas Melton, Klaus H. Kaestner, & Hiroshi Inoue
 Endocrine Pancreas Consortium
 Harvard University, Howard Hughes Medical Institute
 Dept of Molecular and Cellular Biology, 7 Divinity Ave, Cambridge, MA 02138
 Tel: 617-495-1812
 Fax: 617-495-8557
 Email: dmelton@biohp.harvard.edu
 Library was constructed by Dr. Hiroshi Inoue DNA sequencing by: Washington University Genome Sequencing Center For information on obtaining a clone please contact: Dr. Hiroshi Inoue (hinoue@im.wustl.edu)
 Seq primer: -40UP from Gibco
 High quality sequence stop: 361.
 Location/Qualifiers

FEATURES

1. 520
 /organism="Homo sapiens"
 /mol_type="mRNA"
 /db_xref="taxon:9606"
 /clone="IMAGE:6132129"
 /tissue_type="Purified pancreatic islet"
 /lab_host="DH10B"
 /clone_lib="HR85 islet"

/note="Organ: Pancreas; Vector: pBluescript SK(-); Site 1: Not1; Site 2: Xho1; cDNA made by oligo-dT priming. Size-selected on agarose gel. Average insert size ~1kb. 5' Xho1 site was destroyed after directional cloning. Amplified once. Contact information: Hiroshi Inoue, MD, Metabolism Div. (Alan Permutt Lab), Washington University School of Medicine, Box 8127, 660 South Euclid Ave., St. Louis, MO 63110, E-mail: hinoue@imgate.wustl.edu, Tel: 314-362-1916, Fax: 314-747-2692."

ORIGIN

Query Match 18.7%; Score 130.8; DB 5; Length 520;
 Best Local Similarity 94.3%; Pred. No. 2.1e-19;
 Matches 132; Conservative 3; Mismatches 5; Indels 0; Gaps 0;
 QY 221 ACTTCCAGGTGCTCCAGACACATGACAACTGCTAYGACAGGCCAAGAGCTGGACAG 280
 |||
 Db 335 ACTGGACAAGTGTCTCCAGACACATGACAACTGCTATGACAGGCCAAGAGCTGGACAG 276
 |||
 QY 281 CTGTAAATTTCTGCTGGACAMMCCGTACACCCACACCTATTTCATCATCTGCTCTGGCTC 340
 |||
 Db 275 CTGTAAATTTCTGCTGGACAMCCCGTACACCCACACCTATTTCATCATCTGCTCTGGCTC 216
 |||
 QY 341 GGCAATCACCTGTAGCAGTA 360
 |||
 Db 215 GGCAATCACCTGTAGCAGCA 196
 |||

RESULT 8

AW583679
 LOCUS
 DEFINITION
 ia03a06.y1 Human Pancreatic Islets Homo sapiens cDNA clone IMAGE:5637202 5', similar to gb:M21054 PHOSPHOLIPASE A2 PRECURSOR (HUMAN); mRNA sequence.
 AW583679
 ACCESSION
 AW583679.1 GI:7260553
 VERSION
 EST.
 KEYWORDS
 SOURCE
 ORGANISM
 Homo sapiens (human)
 Homo sapiens
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini; Hominidae; Homo.
 REFERENCE
 1 (bases 1 to 522)

AUTHORS

Melton, D., Meadows, A., Clifton, S., Hillier, L., Marra, M., Pape, D., Wyllie, T., Martin, J., Blistain, A., Schmitt, A., Theising, B., Ritter, E., Ronko, I., Bennett, J., Cardenas, M., Gibbons, M., McCann, R., Cole, R., Tsagarishvili, R., Williams, T., Jackson, Y., and Bowers, Y.
 WashU-Harvard Pancreas EST Project
 Unpublished (2000)
 Other ESTs: ia03a06.x3
 Contact: Douglas Melton, Klaus H. Kaestner, & Hiroshi Inoue
 Endocrine Pancreas Consortium
 Harvard University, Howard Hughes Medical Institute
 Dept of Molecular and Cellular Biology, 7 Divinity Ave, Cambridge, MA 02138
 Tel: 617-495-1812
 Fax: 617-495-8557
 Email: dmelton@biohp.harvard.edu
 Libraries were constructed by Dr. Douglas Melton
 DNA sequencing by: Washington University Genome Sequencing Center For information on obtaining a clone please contact: Juliana Brown (brown@fas.harvard.edu)
 This sequence now available from the IMAGE consortium, for clone orders contact: info@image.llnl.gov
 Seq primer: -40RP from Gibco
 High quality sequence stop: 397.
 Location/Qualifiers

TITLE
JOURNAL
COMMENT

1. 522
 /organism="Homo sapiens"
 /mol_type="mRNA"
 /db_xref="taxon:9606"
 /clone="IMAGE:5637202"
 /tissue_type="Islets of Langerhans"
 /dev_stage="Adult"
 /lab_host="DH10B"
 /clone_lib="Human Pancreatic Islets"
 /note="Organ: Pancreas; Vector: pSPORT1; Site 1: Not 1; Site 2: Sal 1; Library constructed using SuperScript Plasmid Library kit (Life Technologies). cDNA made by oligo-dT priming. Size-selected by column fractionation; average insert size 1.08 kb. Primary library, unamplified."

FEATURES

source
 1. 522
 /organism="Homo sapiens"
 /mol_type="mRNA"
 /db_xref="taxon:9606"
 /clone="IMAGE:5637202"
 /tissue_type="Islets of Langerhans"
 /dev_stage="Adult"
 /lab_host="DH10B"
 /clone_lib="Human Pancreatic Islets"
 /note="Organ: Pancreas; Vector: pSPORT1; Site 1: Not 1; Site 2: Sal 1; Library constructed using SuperScript Plasmid Library kit (Life Technologies). cDNA made by oligo-dT priming. Size-selected by column fractionation; average insert size 1.08 kb. Primary library, unamplified."

ORIGIN

Query Match 18.7%; Score 130.8; DB 1; Length 522;
 Best Local Similarity 94.3%; Pred. No. 2.1e-19;
 Matches 132; Conservative 3; Mismatches 5; Indels 0; Gaps 0;
 QY 221 ACTTCCAGGTGCTCCAGACACATGACAACTGCTAYGACAGGCCAAGAGCTGGACAG 280
 |||
 Db 192 ACTGGACAAGTGTCTCCAGACACATGACAACTGCTATGACAGGCCAAGAGCTGGACAG 251
 |||
 QY 281 CTGTAAATTTCTGCTGGACAMMCCGTACACCCACACCTATTTCATCATCTGCTCTGGCTC 340
 |||
 Db 252 CTGTAAATTTCTGCTGGACAMCCCGTACACCCACACCTATTTCATCATCTGCTCTGGCTC 311
 |||
 QY 341 GGCAATCACCTGTAGCAGTA 360
 |||
 Db 312 GGCAATCACCTGTAGCAGCA 331
 |||

RESULT 9

BQ271518/c
 LOCUS
 DEFINITION
 BQ271518 526 bp mRNA linear EST 15-JUL-2003
 ik15a01.x1 HR85 islet Homo sapiens cDNA clone IMAGE:5781000 3', similar to SW:PA21_HUMAN P04054 PHOSPHOLIPASE A2 PRECURSOR ; mRNA sequence.
 BQ271518
 ACCESSION
 BQ271518.1 GI:20496584
 VERSION
 EST.
 KEYWORDS
 SOURCE
 ORGANISM
 Homo sapiens (human)
 Homo sapiens
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini; Hominidae; Homo.

Hillier, L., Marra, M., Pope, D., Wylie, T., Martin, J., Blistain, A., Schmitt, A., Theising, B., Ritter, E., Ronko, I., Bennett, J., Cardenas, M., Gibbons, M., McCann, R., Cole, R., Teagareishvili, R., Williams, T., Jackson, Y., and Bowers, Y.
Endocrine Pancreas Consortium
Unpublished (2000)
Other ESTs: fj24n05.x1
Contact: Douglas Melton, Klaus H. Kaestner, & Hiroshi Inoue
Endocrine Pancreas Consortium
Harvard University, Howard Hughes Medical Institute
Dept of Molecular and Cellular Biology, 7 Divinity Ave, Cambridge, MA 02138
Tel: 617-495-1812
Fax: 617-495-8557
Email: dmelton@biohp.harvard.edu
Library was constructed by Dr. Douglas Melton DNA sequencing by: Washington University Genome Sequencing Center. This clone is available royalty-free through LNL; please contact the IMAGE Consortium (info@image.llnl.gov) for further information
Putative full length read
vector to vector length is 572
Seq primer: -40RP from Gibco
High quality sequence stop: 475.
Location/Qualifiers

FEATURES

1..529
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/clone="IMAGE:6135800"
/sex="Both"
/tissue_type="Islets of Langerhans"
/dev_stage="Adult"
/lab_host="DH10B"
/clone_lib="Melton Normalized Human Islet 4 N4-HIS 1"
/note="Organ: Pancreas; Vector: pSPORT1; Site1: Not 1; Site2: Sal 1; Starting library constructed using SuperScript Plasmid Library kit (Life Technologies). cDNA made by oligo-dr priming. Size-selected by column fractionation; average insert size 1.08 kb. Library was amplified once on solid support and plasmid DNA from library was prepared. The library DNA was normalized by method #4 from Bonaldo, Lennon, and Soares 1996 Genome Research 6:791-806; 0.5 microgram single-stranded library plasmid DNA was mixed with 5 micrograms PCR product representing library inserts and hybridized to an EcoT of 20. Single-stranded (unhybridized) plasmids were isolated by hydroxyapatite chromatography and used to make this library."

ORIGIN

Query Match 18.7%; Score 130.8; DB 3; Length 529;
Best Local Similarity 94.3%; Pred. No. 2.1e-19;
Matches 132; Conservative 3; Mismatches 5; Indels 0; Gaps 0;
Qy 221 ACTTTCAGGTGCTGCCAGACACATGCAACTGCTAYGACCGCCAGAGCTGGACAG 280
Db 190 ACTGGACAAAGTGTGCCAGACACATGCAACTGCTATGACCGCCAGAGCTGGACAG 249
Qy 281 CTGTAAATTTCTGCTGGACACACCGGTACACCCACACCTATTTCATCTGCTCTGGCTC 340
Db 250 CTGTAAATTTCTGCTGGACACACCGGTACACCCACACCTATTTCATCTGCTCTGGCTC 309
Qy 341 GGCAATCACCTGTAGCAGTA 360
Db 310 GGCAATCACCTGTAGCAGCA 329

RESULT 12

AA835315/c
LOCUS
DEFINITION
ak66e08.s1 Barstead pancreas HPLRBL Homo sapiens cDNA clone
IMAGE:1412870 3' similar to gb:M21054 PHOSPHOLIPASE A2 PRECURSOR
(HUMAN); mRNA sequence.
AA835315 532 bp mRNA linear EST 23-FEB-1998
ak66e08.s1 Barstead pancreas HPLRBL Homo sapiens cDNA clone
IMAGE:1412870 3' similar to gb:M21054 PHOSPHOLIPASE A2 PRECURSOR
(HUMAN); mRNA sequence.

ACCESSION
VERSION
KEYWORDS
SOURCE
ORGANISM

AA835315
AA835315.1 GI:2909043
EST.
Homo sapiens (human)
Homo sapiens

REFERENCE
AUTHORS

1 (bases 1 to 532)
Hillier, L., Allen, M., Bowles, L., Dubuque, T., Geisel, G., Jost, S., Krizman, D., Kucaba, T., Lacy, M., Lennon, G., Marra, M., Martin, J., Moore, B., Schellenberg, K., Steptoe, M., Tan, F., Theising, B., White, Y., Wylie, T., Waterston, R. and Wilson, R.
WashU-NCI human EST Project
Unpublished (1997)
Contact: Wilson RK
Washington University School of Medicine
4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108
Tel: 314 286 1800
Fax: 314 286 1810
Email: est@watson.wustl.edu
This clone is available royalty-free through LNL; contact the IMAGE Consortium (info@image.llnl.gov) for further information.
Seq primer: -40ml3 fwd. ET from Amersham
High quality sequence stop: 458.
Location/Qualifiers

TITLE
JOURNAL
COMMENT

1..532
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/clone="IMAGE:1412870"
/sex="female"
/dev_stage="adult, 34 years"
/lab_host="DH10B"
/clone_lib="Barstead pancreas HPLRBL"
/note="Organ: pancreas; Vector: pT7T3D-Pac (Pharmacia) with a modified polylinker; Site 1: EcoRI; Site 2: NotI; 1st strand cDNA was primed with a Not I - oligo(dT) primer [5', TGTTAGCAATCTCAAGTGGAGCGCGCCCTTTTTTTTTTTTTTTTTTTTTT 3']; double-stranded cDNA was ligated to Eco RI adaptors [AATTCGGATCCCTG], digested with Not I and cloned into the Not I and Eco RI sites of the modified pT7T3 vector. Library constructed by Bob Barstead."

FEATURES

1..532
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/clone="IMAGE:1412870"
/sex="female"
/dev_stage="adult, 34 years"
/lab_host="DH10B"
/clone_lib="Barstead pancreas HPLRBL"
/note="Organ: pancreas; Vector: pT7T3D-Pac (Pharmacia) with a modified polylinker; Site 1: EcoRI; Site 2: NotI; 1st strand cDNA was primed with a Not I - oligo(dT) primer [5', TGTTAGCAATCTCAAGTGGAGCGCGCCCTTTTTTTTTTTTTTTTTTTTTT 3']; double-stranded cDNA was ligated to Eco RI adaptors [AATTCGGATCCCTG], digested with Not I and cloned into the Not I and Eco RI sites of the modified pT7T3 vector. Library constructed by Bob Barstead."

ORIGIN

Query Match 18.7%; Score 130.8; DB 1; Length 532;
Best Local Similarity 94.3%; Pred. No. 2.1e-19;
Matches 132; Conservative 3; Mismatches 5; Indels 0; Gaps 0;
Qy 221 ACTTTCAGGTGCTGCCAGACACATGCAACTGCTAYGACCGCCAGAGCTGGACAG 280
Db 344 ACTGGACAAAGTGTGCCAGACACATGCAACTGCTATGACCGCCAGAGCTGGACAG 285
Qy 281 CTGTAAATTTCTGCTGGACACACCGGTACACCCACACCTATTTCATCTGCTCTGGCTC 340
Db 284 CTGTAAATTTCTGCTGGACACACCGGTACACCCACACCTATTTCATCTGCTCTGGCTC 225
Qy 341 GGCAATCACCTGTAGCAGTA 360
Db 224 GGCAATCACCTGTAGCAGCA 205

RESULT 13

BU950978
LOCUS
DEFINITION
BU950978
BU950978.1 GI:24202730
EST.
Homo sapiens (human)
BU950978 537 bp mRNA linear EST 21-OCT-2002
io74d05.y1 HR85 islet Homo sapiens cDNA clone IMAGE:6132129 5',
similar to SW:PA21_HUMAN P04054 PHOSPHOLIPASE A2 PRECURSOR ; mRNA
sequence.
BU950978
BU950978.1 GI:24202730
EST.
Homo sapiens (human)

Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini;
Hominidae; Homo.

REFERENCE AUTHORS

1 (bases 1 to 552)
Melton, D., Brown, J., Kenty, G., Permutt, A., Lee, C., Kaestner, K.,
Lemishka, I., Scarce, M., Brestelli, J., Gradwohl, G., Clifton, S.,
Hillier, L., Narra, M., Pape, D., Wylie, F., Martin, J., Blistain, A.,
Schmitt, A., Theising, B., Ritter, E., Ronko, I., Bennett, J.,
Cardenas, M., Gibbons, M., McCann, R., Cole, R., Teagareishvili, R.,
Williams, T., Jackson, Y. and Bowers, Y.

TITLE JOURNAL COMMENT

Endocrine Pancreas Consortium
Unpublished (2000)
Contact: Douglas Melton, Klaus H. Kaestner, & Hiroshi Inoue
Endocrine Pancreas Consortium
Harvard University, Howard Hughes Medical Institute
Dept of Molecular and Cellular Biology, 7 Divinity Ave, Cambridge,
MA 02138

Tel: 617-495-1812
Fax: 617-495-8557

Email: dmelton@biohp.harvard.edu

Library was constructed by Dr. Hiroshi Inoue DNA sequencing by:
Washington University Genome Sequencing Center For information on
obtaining a clone please contact: Dr. Hiroshi Inoue
(hinoue@im.wustl.edu)

Seq primer: -40RP from Gibco

High quality sequence stop: 473.

FEATURES source

1..552
Location/Qualifiers
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/tissue_type="Purified pancreatic islet"
/lab_host="DH10B"
/clone_lib="HR85 islet"
/note="Organ: Pancreas; Vector: pBluescript SK(-); Site_1:
NotI; Site_2: XhoI; cDNA made by oligo-dr priming.
Size-selected on agarose gel. Average insert size ~1kb. 5'
XhoI site was destroyed after directional cloning.
Amplified once. Contact information: Hiroshi Inoue, MD,
Metabolism Div. (Alan Permutt Lab), Washington University
School of Medicine, Box 8127, 660 South Euclid Ave., St.
Louis, MO 63110, E-mail: hinoue@imgate.wustl.edu, Tel:
314-362-1916, Fax: 314-747-2692."

ORIGIN

Query Match	18.7%	Score 130.8;	DB 3;	Length 552;
Best Local Similarity	94.3%	Pred. No. 2.1e-19;		
Matches 132;	Conservative 3;	Mismatches 5;	Indels 0;	Gaps 0;

Qy	221	ACTTTCACGGTGTGCCAGACACATGCAACTGTCTAYGACCGCCCAAGAGCTGGACAG	280
Db	201	ACTGGACAAGTGTCTGCCAGACACATGCAACTGTCTATGACCGCCCAAGAGCTGGACAG	260
Qy	281	CTGTAAATTTCTGTGGACAMMCGGTACACCCACCTATTTCATCTGCTCTGGCTC	340
Db	261	CTGTAAATTTCTGTGGACAACCGGTACACCCACCTATTTCATCTGCTCTGGCTC	320
Qy	341	GGCAATCACCTGTAGCGTA	360
Db	321	GGCAATCACCTGTAGCGCA	340

Search completed: February 10, 2006, 00:58:05
Job time : 2623.65 secs

THIS PAGE BLANK (USPTO)

GenCore version 5.1.7
Copyright (c) 1993 - 2006 Bioceleration Ltd.

OM nucleic - nucleic search, using sw model

Run on: February 9, 2006, 14:19:56 ; Search time 1611.05 Seconds
(without alignments)
10346.286 Million cell updates/sec

Title: US-10-607-806-1-G7328-T9182_COPY_7000_9500

Perfect score: 2499.4

Sequence: 1 gtcgtgtcactgtgtccag.....tttgagaccagcctggacaa 2501

Scoring table: IDENTITY NUC
Gapop 10.0 , Gapext 1.0

Searched: 4996997 seqs, 3332346308 residues

Total number of hits satisfying chosen parameters: 9993994

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : N_Geneseq_21.*

- 1: Geneseqn1980s.*
- 2: Geneseqn1990s.*
- 3: Geneseqn2000s.*
- 4: Geneseqn2001as.*
- 5: Geneseqn2001bs.*
- 6: Geneseqn2002as.*
- 7: Geneseqn2002bs.*
- 8: Geneseqn2003as.*
- 9: Geneseqn2003bs.*
- 10: Geneseqn2003cs.*
- 11: Geneseqn2003ds.*
- 12: Geneseqn2004as.*
- 13: Geneseqn2004bs.*
- 14: Geneseqn2005s.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	2499.4	100.0	12174	12	ADJ09983
2	2499.4	100.0	13612	6	ABK47376 Human pho
3	2498.6	100.0	12174	12	ADJ35082 Human PLA
4	457	18.3	42360	13	ABD33466 Human can
5	434.2	17.4	93544	13	ABD33504 Human can
6	434.2	17.4	160482	11	ACN43914 Human gen
7	433.2	17.3	91760	11	ACN44410 Human gen
8	425.4	17.0	38538	13	ABD33357 Human can
9	417.2	16.7	122888	6	ABK83569 Human cdn
10	416.2	16.7	6405	5	AAF97850 Human neu
11	416	16.6	126990	12	ADP13332 Human cel
12	414.2	16.6	196686	11	ACN44170 Human gen
13	413.6	16.5	86000	12	ADP68568 Human PPA
14	413.6	16.5	215974	12	ADQ97523 Human can
15	413.4	16.5	13670	6	AAL42369 Human Gua
16	413.2	16.5	6519	5	ABAI6909 Human ner
17	408.8	16.4	17245	4	AAK83897 Human imm
18	406.2	16.3	348101	12	ADQ97146 Human can
19	405.8	16.2	227246	13	ABD33272 Human can

C 20	404	16.2	95240	10	ADL13556
C 21	403.6	16.1	23639	6	ABQ78991
C 22	400.4	16.0	14176	4	AAS26670 Human gen
C 23	400.4	16.0	14176	8	ABX74019 Human nov
C 24	400.4	16.0	177531	8	ACF62732 Cancer ba
C 25	400.4	16.0	177531	8	ADB20847 MRPI base
C 26	400.4	16.0	177531	10	ADB87936 Human UGT
C 27	400.4	16.0	177531	10	ADB87936 Human MDR
C 28	400.4	16.0	177531	10	ADB92110 Human MDR
C 29	400.4	16.0	177531	10	ADH74617 Human BAC
C 30	399	16.0	87687	11	ACN45166 Human gen
C 31	399	16.0	131078	14	ADX06911 Cyclin-de
C 32	398.6	15.9	174448	11	ACN43946 Human gen
C 33	398.4	15.9	167343	6	ABL64403 Stomach c
C 34	398.4	15.9	167343	6	ABL67239 Thyroid c
C 35	394.8	15.8	30620	4	AAK66931 Human imm
C 36	394.6	15.8	19300	12	ADP74371 Human X c
C 37	392.4	15.7	9289	10	ADE84033 5' regula
C 38	391.4	15.7	226475	9	AAS48279 Human tum
C 39	390	15.6	13224	4	AAS41751 Genomic s
C 40	390	15.6	13224	4	ABA06811 Human gen
C 41	390	15.6	13224	6	ABV84148 Human pol
C 42	390	15.6	36221	4	AAS00624 Human dea
C 43	389.6	15.6	23456	13	ABD33110 Human can
C 44	389.6	15.6	60057	11	ACN44314 Human gen
C 45	388	15.5	75252	11	ACN44450 Human gen

ALIGNMENTS

RESULT 1

ADJ09983

ID ADJ09983 standard; DNA; 12174 BP.

XX AC ADJ09983;

XX 17-JUN-2004 (first entry)

XX Human phospholipase A2 (PLA2G1B) DNA SeqID 1.

XX human; gene; ds; fat reduction; fat deposition; phospholipase A2;
XX PLA2G1B; chromosome 12q24; single nucleotide polymorphism; SNP;
XX appetite suppressant; lipase inhibitor; exercise regimen; obesity;
XX non-insulin dependent diabetes mellitus; NIDDM; cardiovascular disorder;
XX hypertension; antidiabetic.

XX Homo sapiens.

Key	Location/Qualifiers
FT variation	/tag= a
FT variation	/standard name= "Single nucleotide polymorphisms"
FT variation	replace(839,a)
FT variation	/tag= b
FT variation	/standard name= "Single nucleotide polymorphisms"
FT variation	replace(4050,a)
FT variation	/tag= c
FT variation	/standard name= "Single nucleotide polymorphisms"
FT variation	replace(4689,t)
FT variation	/tag= d
FT variation	/standard name= "Single nucleotide polymorphisms"
FT variation	replace(6282,a)
FT variation	/tag= e
FT variation	/standard name= "Single nucleotide polymorphisms"
FT variation	replace(6358,c)
FT variation	/tag= f
FT variation	/standard name= "Single nucleotide polymorphisms"
FT variation	replace(6653,t)
FT variation	/tag= g
FT variation	/standard name= "Single nucleotide polymorphisms"
FT variation	replace(7256,t)
FT variation	/tag= h

FT variation /standard_name= "Single nucleotide polymorphisms"
 FT replace(7300,a)
 FT /*tag= i
 FT variation /standard_name= "Single nucleotide polymorphisms"
 FT replace(7301,c)
 FT /*tag= j
 FT variation /standard_name= "Single nucleotide polymorphisms"
 FT replace(7328,g)
 FT /*tag= k
 FT variation /standard_name= "Single nucleotide polymorphisms"
 FT replace(8062,c)
 FT /*tag= l
 FT variation /standard_name= "Single nucleotide polymorphisms"
 FT replace(9182,t)
 FT /*tag= m
 FT variation /standard_name= "Single nucleotide polymorphisms"
 FT replace(11649,c)
 FT /*tag= n
 FT /standard_name= "Single nucleotide polymorphisms"
 XX WO2004002296-A2.
 XX 08-JAN-2004.
 XX 27-JUN-2003; 2003WO-US020831.
 XX 27-JUN-2002; 2002US-0392362P.
 XX (SEQU-) SEQUENOM INC.
 XX Adam GIR, Langdown ML, Denissenko MF, Dennis E, Cantor C;
 XX Rubin B;
 XX WPI; 2004-071944/07.
 XX P-PSDB; ADJ09984.
 XX Identifying a candidate therapeutic for fat reduction, useful for
 XX treating diabetes, by introducing a test molecule to a system comprising
 XX PLA2G1B protein or nucleic acid, and determining the presence of
 XX interaction between the compounds.
 XX Claim 1; SEQ ID NO 1; 116pp; English.
 XX This invention relates to a novel candidate therapeutic agent useful for
 XX fat reduction and disorders related to fat depositions. Specifically, it
 XX refers to polymorphic variations in the phospholipase A2 (PLA2G1B) DNA,
 XX which is located on chromosome 12q24 and has been associated with central
 XX fat deposition. The present invention describes methods to detect the
 XX presence or absence of these single nucleotide polymorphisms of PLA2G1B,
 XX in particular G7328A and T9182G, and subsequently provide treatment that
 XX reduces fat deposition. This treatment may consist of an appetite
 XX suppressant, a lipase inhibitor, a phospholipase inhibitor, an exercise
 XX regimen, a dietary regimen, psychological counseling, psychotherapy or a
 XX psychotherapeutic. Accordingly, PLA2G1B is a target for reducing fat
 XX deposition and it can be used to treat both obesity and non-insulin
 XX dependent diabetes mellitus (NIDDM), as well as cardiovascular disorders
 XX such as hypertension. As such, it exhibits antidiabetic activity. This
 XX polymucleotide sequence is the human PLA2G1B DNA of the invention.
 XX Sequence 12174 BP; 3220 A; 2996 C; 2739 G; 3219 T; 0 U; 0 Other;
 SQ Query Match 100.0%; Score 2499.4; DB 12; Length 12174;
 Best Local Similarity 99.8%; Pred. No. 0;
 Matches 2497; Conservative 4; Mismatches 0; Gaps 0;
 Qy 1 GTCGTGTCACCTGCTGCCAGCTGGTAAACAGAGCACTCTGTCTCAAAAAAATG 60
 Db 7000 GTCGTGTCACCTGCTGCCAGCTGGTAAACAGAGCACTCTGTCTCAAAAAAATG 7059
 Qy 61 CTTTCAATAAATATATGATAAAGCACTTATATTTTCAAGCCATAGGATCATTTCTCC 120
 Db 7060 CTTTCAATAAATATATGATAAAGCACTTATATTTTCAAGCCATAGGATCATTTCTCC 7119

Qy 121 TGAAGCATCTTGGCGAAGTCATCCCACTGTTTCTGAGAGTGGGAGGTGAGGGCTGAC 180
 Db 7120 TGAAGCATCTTGGCGAAGTCATCCCACTGTTTCTGAGAGTGGGAGGTGAGGGCTGAC 7179
 Qy 181 CTATTGCTCTGCACTTACTCTATCTCAGTGTCCCTCCACATTTTCCAGGTGTGCAGGA 240
 Db 7180 CTATTGCTCTGCACTTACTCTATCTCAGTGTCCCTCCACATTTTCCAGGTGTGCAGGA 7239
 Qy 241 CACATGACACCTGCTAYGACCGAGCCCAAGAGCTGGACAGCTGTAATTTCTGCTGACA 300
 Db 7240 CACATGACACCTGCTAYGACCGAGCCCAAGAGCTGGACAGCTGTAATTTCTGCTGACA 7299
 Qy 301 MMCGGTACACCCACACACTATTATCTACTCTGCTCGCAATCACCCTGTAGCAGTA 360
 Db 7300 ACCGTACACCCACACACTATTATCTGCTCTGGCTCGGCAATCACCCTGTAGCAGTA 7359
 Qy 361 GGTATTATCCCTTCTTGACCTATGAAATTTAGTTGGTTCTCAGTAGGCCGGGGGAAATA 420
 Db 7360 GGTATTATCCCTTCTTGACCTATGAAATTTAGTTGGTTCTCAGTAGGCCGGGGGAAATA 7419
 Qy 421 ATAGTAACACAGCCCATGATTTAGTGTAAATTTTCTGGTCTGGGAGTGTCTCTTTTA 480
 Db 7420 ATAGTAACACAGCCCATGATTTAGTGTAAATTTTCTGGTCTGGGAGTGTCTCTTTTA 7479
 Qy 481 ATCTCAGAAACACACTATGGGATAGGTACAAATTTATCTCCTTAAACAGATAAGAAACT 540
 Db 7480 ATCTCAGAAACACACTATGGGATAGGTACAAATTTATCTCCTTAAACAGATAAGAAACT 7539
 Qy 541 GAGGCTCAGAAAGCTGAGCTATTTGGCCAAAGATCACACAGCTTGTAGTGGTGACATTT 600
 Db 7540 GAGGCTCAGAAAGCTGAGCTATTTGGCCAAAGATCACACAGCTTGTAGTGGTGACATTT 7599
 Qy 601 GGGTTTTTTTTTGTGTGTAGACAGGGCTTGTCTGTCTGTCACCCAGGCATGAGCAC 660
 Db 7600 GGGTTTTTTTTTGTGTGTAGACAGGGCTTGTCTGTCTGTCACCCAGGCATGAGCAC 7659
 Qy 661 AGTGGTGCAACACTAGTCTACTGAGCTCAAACTCTGAGCTCAAGGGATCTGCTGACC 720
 Db 7660 AGTGGTGCAACACTAGTCTACTGAGCTCAAACTCTGAGCTCAAGGGATCTGCTGACC 7719
 Qy 721 TCAGCCTCCCAAGTAGCTGGGACTACGAGCGTGACCAACACCGCTTGTAAATTAATAA 780
 Db 7720 TCAGCCTCCCAAGTAGCTGGGACTACGAGCGTGACCAACACCGCTTGTAAATTAATAA 7779
 Qy 781 ATTTTGTGTAGAGCTGGGTCTTACTGCTGGCAGGCTTCTTAAACTCTGCTGCTT 840
 Db 7780 ATTTTGTGTAGAGCTGGGTCTTACTGCTGGCAGGCTTCTTAAACTCTGCTGCTT 7839
 Qy 841 CAAGCAATCCTCTTACCTTGGCATCCCAAGTGTGGGATACAGGGGTGAGCCACCATG 900
 Db 7840 CAAGCAATCCTCTTACCTTGGCATCCCAAGTGTGGGATACAGGGGTGAGCCACCATG 7899
 Qy 901 TGGCGCTACTTATTTCTTATTTTCAATTTTCCATTTTCAATAGAAATGTAAGATCCACAGAA 960
 Db 7900 TGGCGCTACTTATTTCTTATTTTCAATTTTCCATTTTCAATAGAAATGTAAGATCCACAGAA 7959
 Qy 961 GGATTACTGCTATTTTCTTCTTTTGGAGCAGAGTCTCCTTCAATCAGCTCAAA 1020
 Db 7960 GGATTACTGCTATTTTCTTCTTTTGGAGCAGAGTCTCCTTCAATCAGCTCAAA 8019
 Qy 1021 CCTCCGTTGAGCTCAGTCAACCTCTGCTCTCCGGGTTCAAGYATTTCTCTGCTCAAGC 1080
 Db 8020 CCTCCGTTGAGCTCAGTCAACCTCTGCTCTCCGGGTTCAAGYATTTCTCTGCTCAAGC 8079
 Qy 1081 CTCTGAGTAGTGGAAATTAACAAGCGTGCAACCATGCTTGGCTAATTTTGTATTTT 1140
 Db 8080 CTCTGAGTAGTGGAAATTAACAAGCGTGCAACCATGCTTGGCTAATTTTGTATTTT 8139
 Qy 1141 TAGCAGAGATGGGGTTTACCATGTTGCCAGGCTGGTCTCAAACTCTCAGCTCAAGTG 1200
 Db 8140 TAGCAGAGATGGGGTTTACCATGTTGCCAGGCTGGTCTCAAACTCTCAGCTCAAGTG 8199
 Qy 1201 ATCTGCTGCTCAGTCTCCCAAGTGTGGAAATTAAGCGTGAGTCACTGTGCTGCTG 1260

Db 8200 ATCTGCTGCTCAGTCTCCCAAGTGTGGAATATATAGGCGTGAGTCACTGTGCTGGC 8259
Qy 1261 CGATTACTGCTATTCTTTATTTGCTATATCCCGAGTCTAGAGCAGTGTCTGACATAT 1320
Db 8260 CGATTACTGCTATTCTTTATTTGCTATATCCCGAGTCTAGAGCAGTGTCTGACATAT 8319
Qy 1321 AGTAGGTGCTCAATAAATAAATTTGATGATGACACGCTAGATATAAATAAATTTCTTTCTT 1380
Db 8320 AGTAGGTGCTCAATAAATAAATTTGATGATGACACGCTAGATATAAATAAATTTCTTTCTT 8379
Qy 1381 TTTTAAACAACTCTTGACAACTTTGCGAATAAATAAATAAATTTGCAATTTGCTTTTCA 1440
Db 8380 TTTTAAACAACTCTTGACAACTTTGCGAATAAATAAATAAATTTGCAATTTGCTTTTCA 8439
Qy 1441 CTTATCACCTGTTTATGACTTTTTCATATTCGCTCAAACTTTTATTTGTTACTGTTTCTT 1500
Db 8440 CTTATCACCTGTTTATGACTTTTTCATATTCGCTCAAACTTTTATTTGTTACTGTTTCTT 8499
Qy 1501 ATTCTTACTATTTTGTGCTCACTGAATAATATGCTTTAAATTTGCTTTATATCATCTCTCTGCTC 1560
Db 8500 ATTCTTACTATTTTGTGCTCACTGAATAATATGCTTTAAATTTGCTTTATATCATCTCTCTGCTC 8559
Qy 1561 CACTTTAGAGGCCAAATTTTACAATCTGATGAAGCTATGAACCTCTCTCCCGAGAGAA 1620
Db 8560 CACTTTAGAGGCCAAATTTTACAATCTGATGAAGCTATGAACCTCTCTCCCGAGAGAA 8619
Qy 1621 TACACACACACACACACTCACACAGTTTTTTTTTAAATGTTGCAACTTACACACAGA 1680
Db 8620 TACACACACACACACACTCACACAGTTTTTTTTTAAATGTTGCAACTTACACACAGA 8679
Qy 1681 AACCTGCATTTAGAGGATGTTTGTTCATATTAATTAATAAATAAATCACTGTTGGGACAGTGA 1740
Db 8680 AACCTGCATTTAGAGGATGTTTGTTCATATTAATTAATAAATAAATCACTGTTGGGACAGTGA 8739
Qy 1741 CTCAGGCTGTAAACACAGTACTTTGGAAGTCCAAGTGGGTGGATCACTTTGAGGTGAGA 1800
Db 8740 CTCAGGCTGTAAACACAGTACTTTGGAAGTCCAAGTGGGTGGATCACTTTGAGGTGAGA 8799
Qy 1801 AGTTCGAGACCGCTGCTCAATATGCTGAACCTTATCTCTACTAATAAATAAATAAATTT 1860
Db 8800 AGTTCGAGACCGCTGCTCAATATGCTGAACCTTATCTCTACTAATAAATAAATAAATTT 8859
Qy 1861 AGCTGGGTGTAGTGTGATGATGCTGCTGTAGTCCAGTACTCGGAGGCTGAGGCAAGAGAA 1920
Db 8860 AGCTGGGTGTAGTGTGATGATGCTGCTGTAGTCCAGTACTCGGAGGCTGAGGCAAGAGAA 8919
Qy 1921 TTGCTTGAACCTGGGAGGCGAGGTTGCAGTGCAGGCGAGATCCACCACTGCCTCCAGC 1980
Db 8920 TTGCTTGAACCTGGGAGGCGAGGTTGCAGTGCAGGCGAGATCCACCACTGCCTCCAGC 8979
Qy 1981 CTGGGCGACACAGCGAGACTTATCTCAAAAAATAAATAAATAAATAAATAAAGATCGGAG 2040
Db 8980 CTGGGCGACACAGCGAGACTTATCTCAAAAAATAAATAAATAAATAAATAAAGATCGGAG 9039
Qy 2041 AGAAACAAACTAATAAGATTTCTGGAAGTAAAGCAGAGATACGTAAATTTATATGTAATAA 2100
Db 9040 AGAAACAAACTAATAAGATTTCTGGAAGTAAAGCAGAGATACGTAAATTTATATGTAATAA 9099
Qy 2101 AGTTTAAATGCAATTTAACTGTAACTTTATTTGTTTATTTGTTTATATAAAAGTAAACAAGC 2160
Db 9100 AGTTTAAATGCAATTTAACTGTAACTTTATTTGTTTATTTGTTTATATAAAAGTAAACAAGC 9159
Qy 2161 CAAAGTAAATGCAACTTCAACTCTACATAAATATCTTATTTATGAAAGTGAAGGCAATCT 2220
Db 9160 CAAAGTAAATGCAACTTCAACTCTACATAAATATCTTATTTATGAAAGTGAAGGCAATCT 9219
Qy 2221 ATAATCTTACTACCCAAAGATAACCACTTACATATTTCTCCAGATTTTGGGCGCATACAC 2280
Db 9220 ATAATCTTACTACCCAAAGATAACCACTTACATATTTCTCCAGATTTTGGGCGCATACAC 9279
Qy 2281 TAGCTTTTTTATTTGGGAAAAATTTCCATGTGCGGCAATACCTTAATTTTCTAAATGTCT 2340

Db 9280 TAGCTTTTTTATTTGGGAAAAATTTCCATGTGCGGCAATACCTAAATTTTCTAAATGTCT 9339
Qy 2341 ATGTAGTATTTCCATTTAAGGATGTTCCATAATTTTAAAAATACATGCTTTAAAGTAGAGA 2400
Db 9340 ATGTAGTATTTCCATTTAAGGATGTTCCATAATTTTAAAAATACATGCTTTAAAGTAGAGA 9399
Qy 2401 AACTAGTGTTCGGCATGTTGGCTCAGCTGTATCCAGCACTTTGGGAGGCCGAGGCAAA 2460
Db 9400 AACTAGTGTTCGGCATGTTGGCTCAGCTGTATCCAGCACTTTGGGAGGCCGAGGCAAA 9459
Qy 2461 TGGATCACTTTGAGGTCCGGAGTTTGAGACCCAGCTGGACAA 2501
Db 9460 TGGATCACTTTGAGGTCCGGAGTTTGAGACCCAGCTGGACAA 9500

RESULT 2
ABK47376
ID ABK47376 standard; DNA; 13612 BP.
XX
AC ABK47376;
XX
DT 18-JUN-2002 (first entry)
XX
DE Human Phospholipase A2, group IB (PLA2G1B) gene.
XX
KW Human; ds; gene; SNP; single nucleotide polymorphism; pancreatitis;
KW pancreatic cancer; Phospholipase A2 group IB; PLA2G1B; gene therapy;
KW haplotype; genotype; chromosome 12q23-q24.1; transgenic; drug screening.
XX
OS Homo sapiens.
XX
Key Location/Qualifiers
FT variation replace(3845,A)
FT /tag= a
FT /label= SNP
FT /note= "Single nucleotide polymorphism"
FT variation replace(3968,A)
FT /tag= b
FT /label= SNP
FT /note= "Single nucleotide polymorphism"
FT CDS 4053..9613
FT /tag= c
FT /product= "pla2G1B"
FT exon 4053..4086
FT /tag= d
FT /number= 1
FT intron 4087..5785
FT /tag= e
FT /number= 1
FT exon 5786..5945
FT /tag= f
FT /number= 2
FT intron 5946..6744
FT /tag= g
FT /number= 2
FT variation replace(6060,A)
FT /tag= h
FT /label= SNP
FT /note= "Single nucleotide polymorphism"
FT exon 6745..6873
FT /tag= i
FT /number= 3
FT variation replace(6844,A)
FT /tag= j
FT /label= SNP
FT /note= "Single nucleotide polymorphism"
FT intron 6874..9489
FT /tag= k
FT /number= 3
FT exon 9490..9613
FT /tag= l
FT /number= 4
FT variation replace(9531,A)

FT	/*tag= m	181	CTATTGCTCTGCACCTTACTCTCTATCTCAGCTGTCCCTCCACCTTTCCAGGGTGTGCTGCCAGA	240	
FT	/label= SNP	Db			
FT	/note= "Single nucleotide polymorphism"	6696	CTATTGCTCTGCACCTTACTCTCTATCTCAGCTGTCCCTCCACCTTTCCAGGGTGTGCTGCCAGA	6755	
XX					
PN	WO200212562-A2.	Qy	241	CACATGACAACTGCTATGACCCAGGCAAGAGCTGGACAGCTGTAAATTTCTGCTGGACA	300
XX		Db			
PD	14-FEB-2002.	6756	CACATGACAACTGCTATGACCCAGGCAAGAGCTGGACAGCTGTAAATTTCTGCTGGACA	6815	
XX					
PF	06-AUG-2001; 2001WO-US024663.	Qy	301	MMCCGTACACCCACACCTATTCTCATCTCGTCTCTGGCTCGGCAATCACCCTGTAGCAGTA	360
XX		Db			
PR	04-AUG-2000; 2000US-0223179P.	6816	ACCCGTACACCCACACCTATTCTCATCTCGTCTGCTCTGGCTCGGCAATCACCCTGTAGCAGTA	6875	
XX					
PA	(GENA-) GENAISSANCE PHARM INC.	Qy	361	GGTTTATCCCTTCTTGACCTATGAATTTCTAGTTGGTTCTCAGTAGGCCGGGGGAAATA	420
XX		Db			
PI	Kazemi A, Kliem SE, Koshy B;	6876	GGTTTATCCCTTCTTGACCTATGAATTTCTAGTTGGTTCTCAGTAGGCCGGGGGAAATA	6935	
XX					
DR	WPI; 2002-303982/34.	Qy	421	ATAGTAACAAACAGCCATGATTTAGTGTAAATTTTCTGGTTCTTGGGCAAGTGTCTCTCTTTA	480
XX	P-PSDB; AAU78667.	Db	6936	ATAGTAACAAACAGCCATGATTTAGTGTAAATTTTCTGGTTCTTGGGCAAGTGTCTCTCTTTA	6995
XX					
PT	Novel isolated human Phospholipase A2, Group IB pancreas polynucleotide, for therapeutic purposes, for studying expression and function of the polynucleotide and for expressing the phospholipase protein.	Qy	481	ATCCTCAGAAACAACTATGGGATAGGTACAATTAATCTCACTTAACAGATAGAAAACT	540
XX		Db	6996	ATCCTCAGAAACAACTATGGGATAGGTACAATTAATCTCACTTAACAGATAGAAAACT	7055
PS	Claim 1; Fig 1; 51pp; English.				
XX		Qy	541	GAGGCTCAGAAAGCTGAGCTATTGTCCTCAAGATCACACAGCTTGTAAAGTGTGACAGTTT	600
CC	The invention relates to an isolated human Phospholipase A2, Group IB (pancreas) (PLA2G1B) polynucleotide comprising a sequence which is a polymorphic variant for a reference sequence for the PLA2G1B gene or its polymorphic variant of a reference sequence for the PLA2G1B gene or its fragment, or a polymorphic variant of a reference sequence for a PLA2G1B cDNA or its fragment. Also included are haplotyping/genotyping the PLA2G1B gene of an individual, predicting the haplotype pair for the PLA2G1B gene of an individual, identifying an association between a trait and at least one haplotype or haplotype pair of the PLA2G1B gene, an isolated genotyping oligonucleotide for detecting a polymorphism in the PLA2G1B gene, a recombinant non-human organism transformed or transfected with the PLA2G1B sequence, where the organism expresses a PLA2G1B protein encoded by the first nucleotide sequence or by the polymorphic variant sequence, an isolated polypeptide comprising a sequence which is a polymorphic variant of a reference sequence for the PLA2G1B protein or its fragment, an anti-PLA2G1B monoclonal antibody, screening for drugs targeting PLA2G1B, a computer system for storing and analysing polymorphism data for the PLA2G1B gene and a genome anthology for PLA2G1B gene. The PLA2G1B variant is useful in studying the expression and function of PLA2G1B, and in expressing PLA2G1B protein for use in screening for candidate drugs to treat diseases related to PLA2G1B activity (e.g. pancreatitis and pancreatic cancer) and for therapeutic purposes. The transgenic organism is useful for studying expression of the PLA2G1B isogenes in vivo, for in vivo screening and testing of drugs targeted against PLA2G1B protein, and for testing the efficacy of therapeutic agents and compounds in a biological system. The antibody is useful for studying the effect of the variation on the biological activity of PLA2G1B as well as on the binding affinity of candidate drugs targeting PLA2G1B. The present sequence is the PLA2G1B gene which is located on chromosome 12q23-q24.1	Qy	7056	GAGGCTCAGAAAGCTGAGCTATTGTCCTCAAGATCACACAGCTTGTAAAGTGTGACAGTTT	7115
CC		Db			
CC		Qy	601	GGGTTTTTTTTTGTGTTTGTAGACAGGGCTTGTCTGTCAACCCAGGCATGAGCAC	660
CC		Db	7116	GGGTTTTTTTTTGTGTTTGTAGACAGGGCTTGTCTGTCAACCCAGGCATGAGCAC	7175
CC		Qy	661	AGTGGTGCAACCATAGGTCACTCGAGCTCAACTCTCTGAGCTCAAGGGATCTGCTGACC	720
CC		Db	7176	AGTGGTGCAACCATAGGTCACTCGAGCTCAACTCTCTGAGCTCAAGGGATCTGCTGACC	7235
CC		Qy	721	TCAGCCTCCCAAGTGTGGGACTACGAGCGTGCACACCACCGCTGGCTTAATTAATAAAA	780
CC		Db	7236	TCAGCCTCCCAAGTGTGGGACTACGAGCGTGCACACCACCGCTGGCTTAATTAATAAAA	7295
CC		Qy	781	ATTTTTTTGTAGACATGGGTCTTACTAGTGTGGCCAGGCTTGTCTTAAACTCTGCGCTT	840
CC		Db	7296	ATTTTTTTGTAGACATGGGTCTTACTAGTGTGGCCAGGCTTGTCTTAAACTCTGCGCTT	7355
CC		Qy	841	CAGCAATCTCTCTACCTTGGGATCCCAAGTGTCTGGATTACAGGGGTGAGCCACCATG	900
CC		Db	7356	CAGCAATCTCTCTACCTTGGGATCCCAAGTGTCTGGATTACAGGGGTGAGCCACCATG	7415
CC		Qy	901	TGCGGCTACTTATTTCTTTTACATTTCCATCTTCCAAATAGAAATGTAAGATCCACAGAACAG	960
CC		Db	7416	TGCGGCTACTTATTTCTTTTACATTTCCATCTTCCAAATAGAAATGTAAGATCCACAGAACAG	7475
CC		Qy	961	GGATTACTGCTATTTTCTTCTTCTTTTGTGAGACAGAGTCTCACTTCAATCACTCAA	1020
CC		Db	7476	GGATTACTGCTATTTTCTTCTTCTTTTGTGAGACAGAGTCTCACTTCAATCACTCAA	7535
CC		Qy	1021	CCTCCGTTAGCTCAGTCACTGCAACCTCTGCTCCCGGTTCAAGYGATTTCTCTGCTGAAGC	1080
CC		Db	7536	CCTCCGTTAGCTCAGTCACTGCAACCTCTGCTCCCGGTTCAAGYGATTTCTCTGCTGAAGC	7595
CC		Qy	1081	CTCTGTAGTAGCTGGAATTAACAAGCGTGACACCACCTGCTTGGCTAATTTTGTATTTT	1140
CC		Db	7596	CTCTGTAGTAGCTGGAATTAACAAGCGTGACACCACCTGCTTGGCTAATTTTGTATTTT	7655
CC		Qy	1141	TAGCAGAGATGGGGTTTACCTGTTGCCAGGCTGGTCTCAAACTCCTGACCTCAAGTG	1200
CC		Db	7656	TAGCAGAGATGGGGTTTACCTGTTGCCAGGCTGGTCTCAAACTCCTGACCTCAAGTG	7715
CC		Qy	1201	ATCTGCTGCTCAGTCTCCCAAGTGTGGAATTAAGCGGTGAGTCACTGTGCTCCCTGGC	1260
CC		Db	7716	ATCTGCTGCTCAGTCTCCCAAGTGTGGAATTAAGCGGTGAGTCACTGTGCTCCCTGGC	7775
CC		Qy	1261	CGATTACTGTCTATTTTCTTTTATTTGCTATATTCCTCCACAGATCTAGAGCAGTGTCTGACATAT	1320

Query Match	100.0%;	Score	2499.4;	DB	6;	Length	13612;
Best Local Similarity	99.8%;	Pred. No.	0;				
Matches	2497;	Conservative	4;	Mismatches	0;	Indels	0;
				Gaps	0;		
Qy	1	GTCTGTGTCACGTGCTGCTCAGCTGGTAAACAGAGCAACTCTGTCTCAAAAAAATAATG	60				
Db	6516	GTCTGTGTCACGTGCTCAGCTGGTAAACAGAGCAACTCTGTCTCAAAAAAATAATG	6575				
Qy	61	CTTTCAATAAATATATGATAAAGACCTTATATTTTCAAGCCATAGGATCATTTCTCC	120				
Db	6576	CTTTCAATAAATATATGATAAAGACCTTATATTTTCAAGCCATAGGATCATTTCTCC	6635				
Qy	121	TGAAGCATCTTTGGGCAAGTCATCCCACTCTTCTGAGAGTGGCGGCTGAGGCTGAC	180				
Db	6636	TGAAGCATCTTTGGGCAAGTCATCCCACTCTTCTGAGAGTGGCGGCTGAGGCTGAC	6695				

Db 7776 CGATTACTGCTCTATTTCTTTATTTGCTATATCCCGAGATCTAGAGCAGTGTCTGACATAT 7835
Qy 1321 AGTAGGTGCTCAATAAATAAATTTGATGATGACAGCCTAGATATAAATCTTTCTTTCTT 1380
Db 7836 AGTAGGTGCTCAATAAATAAATTTGATGATGACAGCCTAGATATAAATCTTTCTTTCTT 7895
Qy 1381 TTTTAAACCAATCTTGACAACTTTGACAGAAATAAATAAATCTTTGCAATCTTGCTTTTCA 1440
Db 7896 TTTTAAACCAATCTTGACAACTTTGACAGAAATAAATAAATCTTTGCAATCTTGCTTTTCA 7955
Qy 1441 CTTATCACCTTTGATGATGCTTTTCAATATGCTCTCAAAACCTTTATTTGTTACTGTTTTTC 1500
Db 7956 CTTATCACCTTTGATGATGCTTTTCAATATGCTCTCAAAACCTTTATTTGTTACTGTTTTTC 8015
Qy 1501 ATTGTTACTATTTTATGTCACCTGAATAATATGCTTTAAATTTGCTTTATACATCTCTGCTC 1560
Db 8016 ATTGTTACTATTTTATGTCACCTGAATAATATGCTTTAAATTTGCTTTATACATCTCTGCTC 8075
Qy 1561 CACTTTAGAGCCCAAAATTTTCAAAATCTGATGAAAGCTATGAACCTCTCCCCAGAGAAA 1620
Db 8076 CACTTTAGAGCCCAAAATTTTCAAAATCTGATGAAAGCTATGAACCTCTCCCCAGAGAAA 8135
Qy 1621 TACACACACACACACACTCACACAGTTTTTTTTTAAATGTTTGAACTTAAGACAAGA 1680
Db 8136 TACACACACACACACACTCACACAGTTTTTTTTTAAATGTTTGCAACTAAGACAAGA 8195
Qy 1681 AACCTGCAATAGAGGATGTTTGTTCATATTAATTTAAATAAATCACTGCTGGGCACGTGA 1740
Db 8196 AACCTGCAATAGAGGATGTTTGTTCATATTAATTTAAATAAATCACTGCTGGGCACGTGA 8255
Qy 1741 CTCAGGCTGTAAACACAGTACTTTGGAAGTCCAAAGTGGGTGGATCACTTTGAGTGCAGA 1800
Db 8256 CTCAGGCTGTAAACACAGTACTTTGGAAGTCCAAAGTGGGTGGATCACTTTGAGTGCAGA 8315
Qy 1801 AGTTTCGAGACAGCCTGGTCAATATGTTGAAACCTTATCTCTACTATAAATAACAAAAATT 1860
Db 8316 AGTTTCGAGACAGCCTGGTCAATATGTTGAAACCTTATCTCTACTATAAATAACAAAAATT 8375
Qy 1861 AGCTGGGTGATGATGATCAGTCCCTGATGCTCCAGCTACTCGGAGGCTGAGGCAAGAGAA 1920
Db 8376 AGCTGGGTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 8435
Qy 1921 TTGCTTGAACCTCGGAGGAGAGGTTCAGTGAAGCCGAGATCCACCACTGCCTCCAGC 1980
Db 8436 TTGCTTGAACCTCGGAGGAGAGGTTCAGTGAAGCCGAGATCCACCACTGCCTCCAGC 8495
Qy 1981 CTGGGCGACACAGCGAGACTCTATCTCAAAAAATAAATAAATAAATAAAGGATCGGAG 2040
Db 8496 CTGGGCGACACAGCGAGACTCTATCTCAAAAAATAAATAAATAAATAAAGGATCGGAG 8555
Qy 2041 AGAAACAAACTAATAAGATTCCTGAAGTAAAGGATGAGATACGTAAATTTATATGTAATA 2100
Db 8556 AGAAACAAACTAATAAGATTCCTGAAGTAAAGGATGAGATACGTAAATTTATATGTAATA 8615
Qy 2101 AGTTTAAATGCAATTTAACTGTAATCTTATGTTTATTTTGGTTATATAAAGTAAACAAGC 2160
Db 8616 AGTTTAAATGCAATTTAACTGTAATCTTATGTTTATTTTGGTTATATAAAGTAAACAAGC 8675
Qy 2161 CAAAAGTAATGCAATTTCAAACTCTACATAAATAATCTTATTTAGAAAGTGGAAAGGCATCT 2220
Db 8676 CAAAAGTAATGCAATTTCAAACTCTACATAAATAATCTTATTTAGAAAGTGGAAAGGCATCT 8735
Qy 2221 ATATCTCTACTACCAAGATAACCAAGTATACATATCTCTCAGATTTTGGGGCATACAC 2280
Db 8736 ATATCTCTACTACCAAGATAACCAAGTATACATATCTCTCAGATTTTGGGGCATACAC 8795
Qy 2281 TAGCTTTTTTATTTGGGAAAAATTTCCATGTGCGAGCATACCTAAATTTTCTAAATGCTCT 2340
Db 8796 TAGCTTTTTTATTTGGGAAAAATTTCCATGTGCGAGCATACCTAAATTTTCTAAATGCTCT 8855
Qy 2341 ATGTAGTATTTCCATTTAAGGATGTTTCCATTAATTTTAAATAATACATGCTTTTAAAGTAGAGA 2400

Db 8856 ATGTAGTATTTCCATTTAAGGATGTTCCATAATTTTAAATAATACATGCTTTAAAGTAGAGA 8915
Qy 2401 AACTAGGTTGGGATGCTGCTACGCTGTATCCAGCACTTTTGGGAGCCGAGGCAAA 2460
Db 8916 AACTAGGTTGGGATGCTGCTACGCTGTATCCAGCACTTTTGGGAGCCGAGGCAAA 8975
Qy 2461 TGGATCACTTTGAGGTCGGAGTTTGGACCAAGCCTGGACAA 2501
Db 8976 TGGATCACTTTGAGGTCGGAGTTTGGACCAAGCCTGGACAA 9016
RESULT 3
ADI35082
ID ADI35082-standard; DNA; 12174 BP.
XX
AC ADI35082;
XX
DT 22-APR-2004 (first entry)
XX
DB Human PLA2G1B nucleotide sequence.
XX
KW PLA2G1B; fat deposition; leanness; polymorphism;
KW non-insulin dependent diabetes mellitus; NIDDM; hyperinsulinemia;
KW hypertension; glucose intolerance; dyslipidemia; hypercoagulability;
KW microalbuminuria; human; gene; ds.
XX
OS Homo sapiens.
XX
PN WO2004002295-A2.
XX
PD 08-JAN-2004.
XX
PF 27-JUN-2003; 2003WO-US020830.
XX
PR 27-JUN-2002; 2002US-0392361P.
XX
PA (SEQU-) SEQUENOM INC.
XX
PI Adam GIR, Langdown ML;
XX
DR WPI; 2004-082843/08.
DR P-PSDB; ADI35083.
XX
PT Diagnosing a predisposition to fat deposition or leanness, useful for
PT diagnosing a predisposition to e.g. diabetes or hypertension, comprises
PT detecting the presence of a polymorphism in the PLA2G1B nucleic acid from
PT the subject.
XX
PS Claim 1; SEQ ID NO 1; 91pp; English.
XX
CC The invention relates to diagnosing a predisposition to fat deposition or
CC leanness in a subject comprising detecting the presence or absence of a
CC polymorphic variation associated with fat deposition at a polymorphic
CC site in a PLA2G1B nucleotide sequence in a nucleic acid sample from a
CC subject, where the presence of the polymorphic variation indicates a
CC predisposition to fat deposition in the subject. The polymorphic
CC variation is a guanine at position 7328 or thymine at position 9182 of
CC the present sequence. The method is useful for diagnosing a
CC predisposition to fat deposition or leanness in a subject, and
CC consequently for diagnosing a predisposition to non-insulin dependent
CC diabetes mellitus (NIDDM) in a subject and conditions such as
CC hyperinsulinemia, hypertension, glucose intolerance, dyslipidemia,
CC hypercoagulability, or microalbuminuria, which can lead to early
CC prescription of preventive measures. The present sequence represents a
CC human PLA2G1B nucleotide sequence.
XX
SQ Sequence 12174 BP; 3217 A; 2992 C; 2738 G; 3215 T; 0 U; 12 Other;
Query Match 100.0%; Score 2499.4; DB 12; Length 12174;
Best Local Similarity 99.9%; Pred. No. 0;
Matches 2499; Conservative 2; Mismatches 0; Indels 0; Gaps 0;
Qy 1 GTGCTGCTCACTGCTGCTCCAGCCTGGGTAAACAGAGCAACTCTGTCTCAAAAAAATG 60

Db 7000 GTCTGTCACTGCTCCAGCTGGTAACAGAGCAACTCTGTCTCAAAAAAAAAAATG 7059
Qy 61 CTTTCAATAAATATATGATAAAGGACTTATATTTTCAAGCCATAGGATCATTTCTCC 120
Db 7060 CTTTCAATAAATATATGATAAAGGACTTATATTTTCAAGCCATAGGATCATTTCTCC 7119
Qy 121 TGAAGCATCTTGGCGAAGTCAATCCCACTGTTCTCAGAGTGCGGAGGTGAGGGCTGAC 180
Db 7120 TGAAGCATCTTGGCGAAGTCAATCCCACTGTTCTCAGAGTGCGGAGGTGAGGGCTGAC 7179
Qy 181 CTATTGCTCTGCACTTACTCTATCTCAGCTGTCCCTCCCACTTTCCAGGTGCTGCCAGA 240
Db 7180 CTATTGCTCTGCACTTACTCTATCTCAGCTGTCCCTCCCACTTTCCAGGTGCTGCCAGA 7239
Qy 241 CACATGCAACTGTATYAGCACAGGCCAAGAGCTGGACAGCTGTAATAATTTCTGCTGACAC 300
Db 7240 CACATGCAACTGTATYAGCACAGGCCAAGAGCTGGACAGCTGTAATAATTTCTGCTGACAC 7299
Qy 301 MMCGTACACCCACACCTATTACATCTGCTCTGCTGCGCAATCACCTGTAGCAGTA 360
Db 7300 MMCGTACACCCACACCTATTACATCTGCTCTGCTGCGCAATCACCTGTAGCAGTA 7359
Qy 361 GGTATTATCCCTTCTGACCTATGAATTTCTAGTTGGTTCTCAGTAGGCCGGGGGAAATA 420
Db 7360 GGTATTATCCCTTCTGACCTATGAATTTCTAGTTGGTTCTCAGTAGGCCGGGGGAAATA 7419
Qy 421 ATAGTAACAAGCCATGATTTAGTTGTTAATTTCTTGGTTCTGCGGCAGTGTCTCCTTTA 480
Db 7420 ATAGTAACAAGCCATGATTTAGTTGTTAATTTCTTGGTTCTGCGGCAGTGTCTCCTTTA 7479
Qy 481 ATCTCTCAGAACACACTATGGGATAGGTACAAATTTCTCTCACTTAAACAGATAAGAAACT 540
Db 7480 ATCTCTCAGAACACACTATGGGATAGGTACAAATTTCTCTCACTTAAACAGATAAGAAACT 7539
Qy 541 GAGGCTCAGAAAGGCTGAGCTATTGCCCCAAGATCACACAGCTTGTCTAAGTGGTGAAGTTT 600
Db 7540 GAGGCTCAGAAAGGCTGAGCTATTGCCCCAAGATCACACAGCTTGTCTAAGTGGTGAAGTTT 7599
Qy 601 GGGTTTTTTTTTGTGTTTAGAGACAGGGTCTTGTCTGTCACCCAGGCATGAGCAC 660
Db 7600 GGGTTTTTTTTTGTGTTTAGAGACAGGGTCTTGTCTGTCACCCAGGCATGAGCAC 7659
Qy 661 AGTGGTGCAACATAGTCACTGAGCGCTCAACTCTCTGAGCTCAAGGGATCTGCTGACC 720
Db 7660 AGTGGTGCAACATAGTCACTGAGCGCTCAACTCTCTGAGCTCAAGGGATCTGCTGACC 7719
Qy 721 TCAGCTCCCAAGTAGCTGGGACTACGAGCGTGCAACCAAGCTGGCTGCTAAATTAAGAAA 780
Db 7720 TCAGCTCCCAAGTAGCTGGGACTACGAGCGTGCAACCAAGCTGGCTGCTAAATTAAGAAA 7779
Qy 781 ATTTTTTTGTAGAGACTGGGCTCTTACTAGTTGGCGAGGCTTGTCTTAAACTCTCGGCTT 840
Db 7780 ATTTTTTTGTAGAGACTGGGCTCTTACTAGTTGGCGAGGCTTGTCTTAAACTCTCGGCTT 7839
Qy 841 CAAGCAATCTCTACCTTGGCATCCCAAGTGTGGGATTAACGGGTGAGCCCAATG 900
Db 7840 CAAGCAATCTCTACCTTGGCATCCCAAGTGTGGGATTAACGGGTGAGCCCAATG 7899
Qy 901 TCGGCTACTTATTTCTTACATTCATCTTTCCAAATAGAAATGTAAGATCCACAGAAACAG 960
Db 7900 TCGGCTACTTATTTCTTTTACATTCATCTTTCCAAATAGAAATGTAAGATCCACAGAAACAG 7959
Qy 961 GGATTAATGCTATTTCTCTGTTCTTTTGTAGACAGAGTCTCAGCTTCATCACTCACTCA 1020
Db 7960 GGATTAATGCTATTTCTCTGTTCTTTTGTAGACAGAGTCTCAGCTTCATCACTCACTCA 8019
Qy 1021 CCTCGTTTCACTCACTCACTCTGCTCCCGGTTTCAAGYATTTCTCTGCTCAAGC 1080
Db 8020 CCTCGTTTCACTCACTCACTCTGCTCCCGGTTTCAAGYATTTCTCTGCTCAAGC 8079
Qy 1081 CTCTGAGTAGCTGGAAATACAAAGCGTGCACCAACCATGCTTGGCTAAATTTTGTATTTT 1140

Db 8080 CTCTGAGTAGCTGGAAATACAAAGCGTGACCAACCATGCTTGGCTAAATTTTGTATTTT 8139
Qy 1141 TACGAGAGTGGGGTTTACCATGTTGGCCAGCGTGTCTCAAACTCTCAAGCTCAAGTG 1200
Db 8140 TACGAGAGTGGGGTTTACCATGTTGGCCAGCGTGTCTCAAACTCTCAAGCTCAAGTG 8199
Qy 1201 ATCTGCTGCTCAGTCTCCCAAAGTGTGGAAATATAGGCGTGAGTCACTGTGCTTGGC 1260
Db 8200 ATCTGCTGCTCAGTCTCCCAAAGTGTGGAAATATAGGCGTGAGTCACTGTGCTTGGC 8259
Qy 1261 CGATTAATGCTATTTCTTTTATTGCTATATCCAGATCTAGAGAGTGTCTGACATAT 1320
Db 8260 CGATTAATGCTATTTCTTTTATTGCTATATCCAGATCTAGAGAGTGTCTGACATAT 8319
Qy 1321 AGTAGTGCTCAATAAATAATGATGAATGCAAGCTAGATATAAATCTTTCTTTTCTT 1380
Db 8320 AGTAGTGCTCAATAAATAATGATGAATGCAAGCTAGATATAAATCTTTCTTTTCTT 8379
Qy 1381 TTTTAAAAACAATCTTGACAACTTTGCAGAAATAAATAACAATCTTGCAATCTGCTTTTCA 1440
Db 8380 TTTTAAAAACAATCTTGACAACTTTGCAGAAATAAATAACAATCTTGCAATCTGCTTTTCA 8439
Qy 1441 CTTATCACTTGTATTGACTTTTTCATATTTGCCCTCAAACTTTATTGTTACTGTTTTC 1500
Db 8440 CTTATCACTTGTATTGACTTTTTCATATTTGCCCTCAAACTTTATTGTTACTGTTTTC 8499
Qy 1501 ATTGTTACTATTTTAGTCACTGAATAATATGGCTTAAATTTGCTTATACATCTCTGCTC 1560
Db 8500 ATTGTTACTATTTTAGTCACTGAATAATATGGCTTAAATTTGCTTATACATCTCTGCTC 8559
Qy 1561 CACTTTAGAAAGGCAAAATTTTCAAACTCTGATGAAGCTATGAACCTCTCCCCAGAGAAA 1620
Db 8560 CACTTTAGAAAGGCAAAATTTTCAAACTCTGATGAAGCTATGAACCTCTCCCCAGAGAAA 8619
Qy 1621 TAC 1680
Db 8620 TAC 8679
Qy 1681 AACCTGATTTAGAGGATGTTGTTTCAATTTAAATAAATCTCAGTTGGGACACAGTA 1740
Db 8680 AACCTGATTTAGAGGATGTTGTTTCAATTTAAATAAATCTCAGTTGGGACACAGTA 8739
Qy 1741 CTCAAGCCTCTAAACACACAGTACTTTTGGAACTTCCAAAGTGGGTGGATCACTTTGAGGTGAGA 1800
Db 8740 CTCAAGCCTCTAAACACACAGTACTTTTGGAACTTCCAAAGTGGGTGGATCACTTTGAGGTGAGA 8799
Qy 1801 AGTTGAGACACAGCCTGGTCAATATAGTGAACACCTATCTCTATAAATAATACAAAAATTT 1860
Db 8800 AGTTGAGACACAGCCTGGTCAATATAGTGAACACCTATCTCTATAAATAATACAAAAATTT 8859
Qy 1861 AGCTGGGTGAGTCACTGATGCTGATGCTCCAGCTACTCGGAGGCTGAGGCAAGAGAA 1920
Db 8860 AGCTGGGTGAGTCACTGATGCTGATGCTCCAGCTACTCGGAGGCTGAGGCAAGAGAA 8919
Qy 1921 TTGCTTGAACCTGGGAGGCGAGAGGTTCAGTGAGCCGAGATCCCAACCACTGCACTCCAGC 1980
Db 8920 TTGCTTGAACCTGGGAGGCGAGAGGTTCAGTGAGCCGAGATCCCAACCACTGCACTCCAGC 8979
Qy 1981 CTGGGCGACACAGCGAGACTCTATCTCAAAAAATAAATAAATAAATAAAGGATCGGAG 2040
Db 8980 CTGGGCGACACAGCGAGACTCTATCTCAAAAAATAAATAAATAAATAAAGGATCGGAG 9039
Qy 2041 AGAAACAAAACTAATAAGATTTCTTGAAGGTAAAGCAGAGATACGTAATATATCTAATAA 2100
Db 9040 AGAAACAAAACTAATAAGATTTCTTGAAGGTAAAGCAGAGATACGTAATATATCTAATAA 9099
Qy 2101 AGTTTAAATGCAATTTTAACTGTAATCTTATTTGTTTATTTTGGTTTATAAAAAGTAAACAGC 2160
Db 9100 AGTTTAAATGCAATTTTAACTGTAATCTTATTTGTTTATTTTGGTTTATAAAAAGTAAACAGC 9159
Qy 2161 CAAAAGTAATGCAACTTCAAACTCTACATAAATATCTATATGAAAAGTGAAGGCACTCT 2220
Db 9160 CAAAAGTAATGCAACTTCAAACTCTACATAAATATCTATATGAAAAGTGAAGGCACTCT 9219

QY 2221 ATATCTCTACTACCAAGATAACACAGTTACATATCTCTCAGATTTTGGGGCATACAC 2280
 DB 9220 ATATCTCTACTACCAAGATAACACAGTTACATATCTCTCAGATTTTGGGGCATACAC 9279
 QY 2281 TAGCTTTTATTTGGGAAATTTCCATGTGCGAGGATACCTTAATTTTCTAAATGTCT 2340
 DB 9280 TAGCTTTTATTTGGGAAATTTCCATGTGCGAGGATACCTTAATTTTCTAAATGTCT 9339
 QY 2341 ATGTAGTATTTCCATTTAAGCATGTTCCATAATTTTAAATATACATGCTTTAAAGTAGAGA 2400
 DB 9340 ATGTAGTATTTCCATTTAAGCATGTTCCATAATTTTAAATATACATGCTTTAAAGTAGAGA 9399
 QY 2401 AACTAGGTTGGGCATGTGCTCACGCTGTATCTCCAGCACTTTGGGAGGCCGAGGCAAA 2460
 DB 9400 AACTAGGTTGGGCATGTGCTCACGCTGTATCTCCAGCACTTTGGGAGGCCGAGGCAAA 9459
 QY 2461 TGGATCATTTGAGGTCGGAGTTTGAGACAGCCTGGACAA 2501
 DB 9460 TGGATCATTTGAGGTCGGAGTTTGAGACAGCCTGGACAA 9500

RESULT 4
 ABD33466/c
 ID ABD33466 standard; DNA; 42360 BP.

XX AC ABD33466;
 XX 18-NOV-2004 (first entry)
 XX Human cancer-associated (CA) gene HD07-089.
 XX Human; cancer-associated protein; CAP; cancer-associated gene; CA; gene;
 KW ds; cancer; cytostatic.
 XX Homo sapiens.
 XX WO2004058146-A2.
 XX 15-JUL-2004.
 XX 15-DEC-2003; 2003WO-US040081.
 XX 17-DEC-2002; 2002US-00322281.
 XX (SAGR-) SAGRES DISCOVERY INC.
 XX Morris DW, Malandro MS;
 PI WPI; 2004-499109/47.
 DR Novel human cancer associated protein encoded within open reading frame
 PT of cancer associated gene, useful as targets for diagnosing cancer.
 XX Claim 16; SEQ ID NO 618; 182pp; English.

XX The invention relates to cancer-associated proteins (CAP) and the cancer-
 CC associated (CA) nucleic acids encoding them. The invention also relates
 CC to a method for treating cancers involving administering to a patient an
 CC inhibitor of CAP, and a method of screening for anticancer activity in a
 CC potential drug involving providing a cell that expresses a CA gene,
 CC contacting a tissue sample derived from a cancer cell with an anticancer
 CC drug candidate and monitoring the effect of the anticancer drug candidate
 CC on expression of the CA gene. The CAP proteins are useful for detecting
 CC cancer associated with expression of a CAP protein in a test cell sample
 CC and for screening for a bioactive agent capable of modulating the
 CC activity of a CAP protein. The CA nucleic acids are useful for diagnosing
 CC cancer, involving determining the expression of a CA nucleic acid in a
 CC tissue. This sequence represents a human CA gene of the invention. Note:
 CC The sequence data for this patent did not form part of the printed
 CC specification, but was obtained in electronic format directly from WIPO
 CC at ftp.wipo.int/pub/published_pct_sequences

SQ Sequence 42360 BP; 11698 A; 9658 C; 10445 G; 10559 T; 0 U; 0 Other;
 Query Match 18.3%; Score 457; DB 13; Length 42360;
 Best Local Similarity 100.0%; Pred. No. 4.3e-84;
 Matches 457; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 2045 ACAAACCTAATAAGATTCCTGAAGGTAAGCAGAGATACGTAATTAATGTAATAAGTT 2104
 DB 42360 ACAAACCTAATAAGATTCCTGAAGGTAAGCAGAGATACGTAATTAATGTAATAAGTT 42301
 QY 2105 TAAATGCATTTTAACTGTAACTCTTATGTTTATTTGGTTATAAAAGTAAACCAAGCAAA 2164
 DB 42300 TAAATGCATTTTAACTGTAACTCTTATGTTTATTTGGTTATAAAAGTAAACCAAGCAAA 42241
 QY 2165 AGTAATGCACACTTCAAACTCTACATAAATATCTATTATGTAAGTGAAGGCACTCTATAA 2224
 DB 42240 AGTAATGCACACTTCAAACTCTACATAAATATCTATTATGTAAGTGAAGGCACTCTATAA 42181
 QY 2225 TCCTACTACCCAAAGATAACCAAGTTACATATTCCTCCAGATTTTGGGGCATACACTAGC 2284
 DB 42180 TCCTACTACCCAAAGATAACCAAGTTACATATTCCTCCAGATTTTGGGGCATACACTAGC 42121
 QY 2285 TTTTATTTTGGGAAATTTCCATGTGCGAGGATACCTTAATTTTCTAAATGTCTATGT 2344
 DB 42120 TTTTATTTTGGGAAATTTCCATGTGCGAGGATACCTTAATTTTCTAAATGTCTATGT 42061
 QY 2345 AGTATTCATTTAAGGATGTTCCATAATTTTAAATATACATGCTTTAAAGTAGAGAACT 2404
 DB 42060 AGTATTCATTTAAGGATGTTCCATAATTTTAAATATACATGCTTTAAAGTAGAGAACT 42001
 QY 2405 AGGTTGGGCATGCTCGCTCACGCTGTATCCAGCACTTTGGGAGGCCGAGGCAAAATGGA 2464
 DB 42000 AGGTTGGGCATGCTCGCTCACGCTGTATCCAGCACTTTGGGAGGCCGAGGCAAAATGGA 41941
 QY 2465 TCACITGAGGTCGGAGTTTGAGACAGCCTGGACAA 2501
 DB 41940 TCACITGAGGTCGGAGTTTGAGACAGCCTGGACAA 41904

RESULT 5

ABD33504
 ID ABD33504 standard; DNA; 93544 BP.
 XX AC ABD33504;
 XX 18-NOV-2004 (first entry)
 XX Human cancer-associated (CA) gene HD07-098.
 XX Human; cancer-associated protein; CAP; cancer-associated gene; CA; gene;
 KW ds; cancer; cytostatic.
 XX Homo sapiens.
 XX WO2004058146-A2.
 XX 15-JUL-2004.
 XX 15-DEC-2003; 2003WO-US040081.
 XX 17-DEC-2002; 2002US-00322281.
 XX (SAGR-) SAGRES DISCOVERY INC.
 XX Morris DW, Malandro MS;
 PI WPI; 2004-499109/47.
 PT Novel human cancer associated protein encoded within open reading frame
 PT of cancer associated gene, useful as targets for diagnosing cancer.
 XX Claim 16; SEQ ID NO 676; 182pp; English.

CC The invention relates to cancer-associated proteins (CAP) and the cancer-
CC associated (CA) nucleic acids encoding them. The invention also relates
CC to a method for treating cancers involving administering to a patient an
CC inhibitor of CAP, and a method of screening for anticancer activity in a
CC potential drug involving providing a cell that expresses a CA gene,
CC contacting a tissue sample derived from a cancer cell with an anticancer
CC drug candidate and monitoring the effect of the anticancer drug candidate
CC on expression of the CA gene. The CAP proteins are useful for detecting
CC cancer associated with expression of a CAP protein in a test cell sample
CC and for screening for a bioactive agent capable of modulating the
CC activity of a CAP protein. The CA nucleic acids are useful for diagnosing
CC cancer, involving determining the expression of a CA nucleic acid in a
CC tissue. This sequence represents a human CA gene of the invention. Note:
CC The sequence data for this patent did not form part of the printed
CC specification, but was obtained in electronic format directly from WIPO
CC at ftp.wipo.int/pub/published_pct_sequences
XX
SQ Sequence 93544 BP; 26138 A; 20027 C; 21635 G; 25744 T; 0 U; 0 Other;

Query Match 17.4%; Score 434.2; DB 13; Length 93544;
Best Local Similarity 59.7%; Pred. No. 2.6e-79;
Matches 861; Conservative 1; Mismatches 559; Indels 22; Gaps 7;

QY 604 TTTTCTTTTGTGTTTATAGACAGGCTTCTGCTGTCCACCCAGGCAATGACAGT 663
DB 18099 TTTTCTTTTGTGTTTATAGACAGGCTTCTGCTGTCCACCCAGGCTGAGTGCAGT 18158

QY 664 GGTGCAACCATAGTCTACTGCAGCCTCAACCTCTGAGCTCAAGGGATCTGTCACCTCA 723
DB 18159 GGCACCATCTTGGGTCACTGCAACCTCCACCATGTGGGTTCAAGAGATTCTACTGCTCA 18218

QY 724 GCCTCCCAAGTAGCTGGGACTACGAGCGTGCACACACCGCTGGCTAATTAATAAATT 783
DB 18219 ACCTCCCAAGTAGCTGGGACTACGAGCGTGTGGCCACACACCTTGTCTAATT--TTTGTGT 18276

QY 784 TTTTGTAGAGACTGGGCTTACTACCTGSCCAGGCTGTCTTAACCTCTGCTCTCAA 843
DB 18277 TTTTATAGAGACAGGGTTTGGCCATGTTGCCAGGCTGATCTCAAACTCTGGCTCAA 18336

QY 844 GCATCTCTCTACCTTGGCATCCAAAGTGTGGGATTACAGGGGTGAGCCACCATGTGC 903
DB 18337 CTGATCCGCTGCTCGGCTCCCAAAGTGTGGGATTACAGGTGTGAGCCACCATGCCT 18396

QY 904 GGCTACTTATTTCTTTACATTCATCTTCCATAGAAATGTAGATCCACAGACAGGA 963
DB 18397 GGCCACCAACAGGCTTTATTTACAGGTTGCCTCATTACCTGAAATCTCTCTCCCAAATC 18456

QY 964 TTACTGCTATTTCTTCTCTTTTGTAGACAGAGTCTCACTTCATCACCTCA--A 1020
DB 18457 TTGTTACCAACACAGGAACATGAGATGGAGTCTCGTCCGTCGCGAGTCTGGAGTGC 18516

QY 1021 CCTCCGTTCACTCACTGCAACCTCTGCTCCCGGTTCAAGYATTTCTCTGCTTAAGC 1080
DB 18517 AGTGATCTCAGCTCACTGCAACCTCTGCTCTCGGTTCAAGTGAATCTTCTGCTCAGC 18576

QY 1081 CTCCTGAGTAGCTGGAAATTAACGGTGCACCATGCTTGGCTAATTTTGTATTTT 1140
DB 18577 CTCCCAAGTAGCTGGGACTACAGGCATCGGCCACCCACCTGGGCTAA--TTTGTATTTT 18635

QY 1141 TAGCAGAGATGGGGTTTACCATGCTGTCAGGCTGTCTCAAACTCTGACCTCAAGTG 1200
DB 18636 TAGTAGAGACGGGGTTTACCATGCTGTCAGGCTGGTCTGAACTCTGACCTC--GTG 18693

QY 1201 ATCTGCTGCTCAGTCTCCAAAGTGTGGAATTAATAGGCGTGAAGTCACTGTGCTGGC 1260
DB 18694 ATCCACCGCTCAGCTTCCCAAAGTGTGGGATTACAGGCATGAGCCACCATACCTGGC 18753

QY 1261 CGATTACTGTCTATTTT-----CTTTATGCTATATCCCGAGATCTAGAGCAGTGTCTG 1314
DB 18754 CAGGAACCTAGACAGTCTTTAGAGACATCTTTCTGAAATAGGGCTGGTCCCTCATTTATTC 18813

QY 1315 ACATATAGTAGGTCTCAATAAATAATTTAGTAATGCACAGCCTAGATATATAAATCTTCTT 1374

DB 18814 ATTAATGCTATTTTCTCCAAACCCCGGAACCAATAGATGTGATAGCATGGCATCCCTC 18873
QY 1375 TTTCTTTTAAAAACAATCTTGACAACTTTCGAGAAATAAATACATCTTGTGATCTGCT 1434
DB 18874 ATGATCTGAAGTTACTGTCCGTAACCTGCGGTCCCATTCGATTTGGGATATTGGGGGATGCC 18933
QY 1435 TTTTCACTTATCA--CCTGCTTATGACTTTTTCATATTGCTCCTCAAACTTTATTGTTACTG 1493
DB 18934 TGCAGTGTCTGTGTCAATGCTGCAGTCTGCGGTGCCATCAGTGTGAGTCCCTCT 18993
QY 1494 TTTTTCATTTTACTATTATTAGTCACTGAATAAATAGCTTAATTTGCTTATACATCCT 1553
DB 18994 GGGCTGCGTCCGTTTTTCTCATCTTCTGATTAATCTCAGCACTTTGTACAATACCTGGC 19053
QY 1554 CTGTCTCACTTTTAGAGGCCAAATTTCAAAATCTGATGAAGACTATGAACCTCTCCCC 1613
DB 19054 ACTCAATGAATTAATAAATAATCTGAAGCGGGGTGTGGGGGAGCTTTGTCAATATAGCCCC 19113
QY 1614 AGAGAATAACACACACACACACACTCACACACAGTTTTTTTTTAAATGTTTGCACATAA 1673
DB 19114 ACCTGTTTGGCTTGGACTTGATGAAGCCGTTTGTGCTATGCAATATGTCCATCTCTT 19173
QY 1674 GACAAGAAACCTGCATTTAGAGGATGTTTTCATATTAAATAAAATACTCA----- 1726
DB 19174 TTCTAGACAATGAGTTTACTTTGGATGAATAGGCAGATCTCTTGGTTAAACCATAAAGACAG 19233
QY 1727 GTTGGCAGAGTCACTCAAGCCTGTAAACACAGTACTTTTGGAAAGTCCAAGTGGGTGGAT 1786
DB 19234 GCTGGCGGGTGGCTCACATCTGTAAATCCAGCACTTTGAGAGGCGGAGCGGCGAGAT 19293
QY 1787 CACTTCAGGTGAGAAGTTCGAGACCAAGCTGTGTCATATATGTTGAAACCTCTATCTCTACTA 1846
DB 19294 CACTTCAGGTGAGAAGTTCGATACCAGCTTGGCAACATGTTGAAACCTCTATCTACTA 19353
QY 1847 AAAATACAAAAATTTAGCTGGGTGTAGTGTGATGCTGTGTAGTCCAGCTACTCTCGGAGG 1906
DB 19354 AAAATACAAAAATTTAGCCGAGTGGGTGGCGATGCTGTAACTCTAGCTACTTTGGAGG 19413
QY 1907 CTGAGCAAGAGAAATTTCTTGAACCTGGGAGGACAGAGTTGCACTGAGCGGAGATCCCAC 1966
DB 19414 CTGAGCAGGAGAAATCGCTTGAACCCGGGAGGTGAAGCTTTCACTGAGCGGAGATTGTGC 19473
QY 1967 CACTGCACCTCCAGCCTGGCGCACACAGCGAGACTCTATCTCAAAAAATAAATAAATAA 2026
DB 19474 CATTACACTCCAGCTTGGCAACAGAGCGAGACTGTGTCTCAAAATGAAACCCATAAAG 19533
QY 2027 ATA 2029
DB 19534 ACA 19536

RESULT 6
ACN43914/c
ID ACN43914 standard; DNA; 160482 BP.
XX
AC ACN43914;
XX
DT 18-NOV-2004 (first entry)
XX Human genomic sequence hCG26773.
XX
XX Cytostatic; carcinoma; lymphoma; cancer; human; gene; ss.
XX Homo sapiens.
OS
PN WO2003073826-A2.
XX
PD 12-SEP-2003.
XX
PF 28-FEB-2003; 2003WO-US006235.
XX
PR 01-MAR-2002; 2002US-00087192.
XX

AC ABD33357;
 XX 18-NOV-2004 (first entry)
 XX Human cancer-associated (CA) gene HD07-065.
 XX Human; cancer-associated protein; CAP; cancer-associated gene; CA; gene;
 XX ds; cancer; cytostatic.
 XX Homo sapiens.
 XX W02004058146-A2.
 XX 15-JUL-2004.
 XX 15-DEC-2003; 2003WO-US040081.
 XX 17-DEC-2002; 2002US-00322281.
 XX (SAGR-) SAGRES DISCOVERY INC.
 XX Morris DW, Malandro MS;
 XX WPI; 2004-499109/47.
 XX Novel human cancer associated protein encoded within open reading frame
 XX of cancer associated gene, useful as targets for diagnosing cancer.
 XX Claim 16; SEQ ID NO 448; 182pp; English.
 XX The invention relates to cancer-associated proteins (CAP) and the cancer-
 XX associated (CA) nucleic acids encoding them. The invention also relates
 XX to a method for treating cancers involving administering to a patient an
 XX inhibitor of CAP, and a method of screening for anticancer activity in a
 XX potential drug involving providing a cell that expresses a CA gene,
 XX contacting a tissue sample derived from a cancer cell with an anticancer
 XX drug candidate and monitoring the effect of the anticancer drug candidate
 XX on expression of the CA gene. The CAP proteins are useful for detecting
 XX cancer associated with expression of a CAP protein in a test cell sample
 XX and for screening for a bioactive agent capable of modulating the
 XX activity of a CAP protein. The CA nucleic acids are useful for diagnosing
 XX cancer, involving determining the expression of a CA nucleic acid in a
 XX tissue. This sequence represents a human CA gene of the invention. Note:
 XX The sequence data for this patent did not form part of the printed
 XX specification, but was obtained in electronic format directly from WIPO
 XX at ftp.wipo.int/pub/published_pct_sequences
 XX
 SQ Sequence 38538 BP; 9418 A; 9893 C; 9523 G; 9684 T; 0 U; 20 Other;
 Query Match 17.0%; Score 425.4; DB 13; Length 38538;
 Best Local Similarity 60.0%; Pred. No. 1.4e-77;
 Matches 879; Conservative 1; Mismatches 552; Indels 33; Gaps 9;
 QY 567 CCAAGATCACACAGCTTGTAAAGTGGTGACAGTTGGGTTTTTTTTTTGGTTGGTTAGAG 626
 DB 35304 CAAAAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAA 35245
 QY 627 ACAGGCTTGTCTC-TGTCAACCCAGGATGACAGAGTGGTGCAACCATAGGTCACTGCA 585
 DB 35244 ATGAAGTTTGTCTTGTGTGCCCAGGCTGGAGTGGCAATGGTGTGATCTTGGCTCACAGCA 35185
 QY 686 GCCTCAACCTCTCTGAGCTCAAGGGATCTGCTGACCTCAGCTCCCAAGTAGCTGGGACTA 745
 DB 35184 ACCTCCACCTCCAGGGTTCAGGCGATCTCTCTCAGCTCCTCAGTAGCTGGGATTA 35125
 QY 746 CGAGCTGCACACACAGCTGCTAATTAATAAATAAATAAATAAATAAATAAATAAATAAATAA 805
 DB 35124 CAGGATGCGCCACACACACAGCTAATTTTGA---TTTTAGTAGAGATGGGTTTCT 35068
 QY 806 CTACGTTGGCCAGGCTTGTCTTAACTCTCTGGCTTCAAGCAATCTCTACCTTGGCATC 865
 DB 35067 CCATGTTGGTCAGGCTGGTCTTGAATCTCTGACCTCAGGTGATCCACCTCTTGGCTTC 35008

QY 866 CCAAAGTGTGGGATTACAGGGGTGAGCCACCATGTGCGGTACTTATTTCCTTTTACATTC 925
 DB 35007 CCGAAGTGTGGGATTACAGGGATGAGCCACTGCGGCTGGC-----CTCAGTTAAGGA 34954
 QY 926 CAFCTTTTCCAAATAGAATGTAAAGATCCACAGAACAGGATTTACTGCTATTTTCTTCTTT 985
 DB 34953 CTTTTTACATATATTATTATTATTATTATTATTATTATTATTATTATTATTATTATTATTATT 34894
 QY 986 CTTTTTTTGACAGACAGAGTCTCAGTTTCATCAGCTCAACCTCCGTTTACGTCACGACCTC 1045
 DB 34893 GTTCTCCAGGCTGGAGTGAATGGCACAGACC-----TTGGCTCACTCAACCTC 34845
 QY 1046 TGCCTCCCGGGTTCAAGYGATTCTCTGCTAAGCCTCTCTGAGTAGCTGGAATTAACAAGC 1105
 DB 34844 CGCCTCCCGGGTTCAAGCGATTCTCTGCTTGGCCTCCCAAGTAGCTGGGATTAACAGC 34785
 QY 1106 GTGCACACCATGCTTGGCTAAATTTTGTATTTTGTAGCAGAGATGGGTTTACCATGT 1165
 DB 34784 GCCTGCCATCATGCTCGGCTAA-TTTTTGTATTTTGTAGTAGAGACGGGGTTTCCACATTA 34726
 QY 1166 TGCCAGGCTGGTCTCAAACTCTGACCTCAAGTATCTGCTGCTCAGTCTCCCAAG 1225
 DB 34725 TGCCAGGCTGGTCTTGAAGTCTGACCTCAGGTATCTGCTGCTCGGCTCCCAAG 34666
 QY 1226 TGTGGAATTATAGGCGTGAGTCACTGTGCTGCGCGATTACTGTCTATTTTCTTTATTG 1285
 DB 34665 TGTGGGATGACAGGTGTGAGCCACACACTCGGCTTTATTATTATTTCGAGACAGG 34606
 QY 1286 CTATATCCCAGATCTAGAGCAGTGTCTGACATATATAGTAGGTGCTCAATAAATAATGAT 1345
 DB 34605 GTATCTC-----TCTGTTGCCAGGCTGGAGTGCAGTGGTGCAGTGGTGTGCA 34552
 QY 1346 GAATGACAGCCTAGATATAAATCTTCTTTTCTTTTCTTTTAAACAATCTTGACACTTT 1405
 DB 34551 GCTCAAACTCTGGGCTCCAGCA--ATTCTCTGCTCAGGCTCTCTGGGTAGTGGAA 34494
 QY 1406 GCAGATAAATAACAATCTTGGCATCTGCTTTTCTATCTATCACTTGTATGACTTTTTC 1465
 DB 34493 TACAGGTGCAATATATCATGCTCGCTAAATTTTAAATTTTGTCAAGATGGGTCTCT 34434
 QY 1466 ATATTGCTCTCAACCTTTTATTTGTTACTGTTTCTTCAATGTTTACTATTTTGTACATGA 1525
 DB 34433 CTATATTGCCAGGCTGGTCTCAAACTCTGCGCTCAAGTATCTCCCTTCTGGGCTTCA 34374
 QY 1526 ATATGCTTAAATTTGCTTATACATCTCTGCTCCACTTTTAGAGGGCCAAATTTACAA 1585
 DB 34373 AAAAGTGGGATTAATAAGTATAGACCCACCCTCCGCGCTTCCCTCTCTCTCTTC 34314
 QY 1586 TCTGATGAAGCTATGAACCTCTCTCCAGAGAAATACACACACACACACACTCACAC 1645
 DB 34313 CTTTCTGCTTCT 34254
 QY 1646 ACAGTTTTTTTTTAAATGTTTGAACCTAAGACAAGAAACCTGCAATAGAGGATTTTGTTC 1705
 DB 34253 CTGTTATGTTCTGATTAACCTGATATATTGATATATTATTTCAATAAAGATGCTTCT 34194
 QY 1706 ATA--TTAATTAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATA 1763
 DB 34193 TTCTTTGGTTAAAAAGTATCCGGCCGCGATGGGTTCACGCTTGTAAACCCAGTACT 34134
 QY 1764 TTGGAAGTCCAAAGTGGGTGGATCACTTGAAGTGAAGATTTCCAGACACAGCCTGGTCAAT 1823
 DB 34133 TTGGGAGGCTGTGGTGGGCGGATCACTGAGATCAGGATCAGACAGCCTGGCCAC 34074
 QY 1824 ATGTGGAACCCCTATCTCTACTTAAATAATAAATAAATAAATAAATAAATAAATAAATAA 1883
 DB 34073 ATGTGTAACCCCTATCTCTACTTAAATAATAAATAAATAAATAAATAAATAAATAAATA 34014
 QY 1884 TGTAGTCCAGCTACTCGGGAGGCTGAGGACAGAAATTTGTTGAACCTTGGGAGCAGAG 1943
 DB 34013 TGTAAATCCAGCTACTTGGGAGGCAAGGAGAAATTTGTTGAACCCGAGGAGCAGAG 33954
 QY 1944 GTTGAGTGGAGCGAGATCCCACTGACCTCCAGCCTGGGCGACACACAGCAGAGACTCTA 2003

Qy	1414	AATACAACTTGCATCTCTGCTTTTTCACCTTATCACCTTGTATGACTTTTTCATATGCC	1473
Db	71059	CATGACGACAGGTTTTCTTTTGTGTTTATTCATTATGGAACCTAAGTCCCTAGAACAA	71118
Qy	1474	TCAAACTTTATTTGTTACTGTTTTTTTCATTTGTTTACTATTTTATTTAGTCACTGAATAATATGCC	1533
Db	71119	TGCTGTGAACATAGTGGGTGTTCAATAAATACTGATTTGAATCAATAAAGAAATATATTTTT	71178
Qy	1534	TTAAATTTGCTTATACATCTCTGCTCCACTTTAGNAGCCAAATTTACAATCTGATGA	1593
Db	71179	GGGTTTTTACCCTATCTGATACGTGGACCTCAAACTCATTTCTGCTGTTCAGTATTTCTATTA	71238
Qy	1594	AAGCTATGAACCTCTCCCCAGAGAAATACACACACACACACTCACACACAGATTTT	1653
Db	71239	TTCTGAAGATGAAATATCACAAATAACAATCTCTATTAATGAACATTTGTTTCATAATTTT	71298
Qy	1654	TTTTTAA----TGTTTGCAACTAAGACAAGAAACCTGCATTTAGAGGATGTTTGTTCATAT	1709
Db	71299	TCATGGACATTTTATATACACAGGTAAGAGTTTTCTCAGGGTATATACCTAAGAATGAAC	71358
Qy	1710	TAATTAATAAATAACTCAGTTGGGCACAGTGACTCAAGCCTGTAAACACAGACTTTTGAA	1769
Db	71359	TGTAGTAATAATCTGGCCGACGCGTGGCTCATGCTGTATCCGACACTTTGGGA	71418
Qy	1770	GTCCAAAGTGGGTGATCACTTGAGTGAGAAAGTTTCGAGACAGAGCTCGTCAATATGGTG	1829
Db	71419	GGCCGAGGCAAGTGGATCAC--AAGGTCAGGGGATCAAGACCATCTGGCCAAACATGGTG	71476
Qy	1830	AAACCTTATCTCTACTAAATAATACAAAATTAGCTGGGTGATGATCATGCTCGCTGAGT	1889
Db	71477	AAACCTTGCTCTACTTAAATAATACAAAATTAGCAGGTTGGGGGGCGTGGCTCTGAGT	71536
Qy	1890	CCAGCTACTCGGGAGGCTGAGGCAAGAGAATTTGCTTGAACCTGGGAGGCGAGAGTTGCA	1949
Db	71537	CCCAGCTACTCAGGAGGCTGAGCGAGGAAATCGTTTGAAACCCAGAGGCTGGAGGTTGCA	71596
Qy	1950	GTGAGCGGAGATCCCACTTCCATCTCAGCTGGGGCGACACAGCGAGACTCTATCTCAA	2009
Db	71597	GTGAGCGAAGTCCGACCACTGCACCTCCAGCCTGGGTGACAGACGACTCCATCTCAC	71656
Qy	2010	AAAAATAAATAAATAAATAA	2031
Db	71657	AAACCAAAAGAAAAA	71678

RESULT 10

RESULI IO
AAE97850/C

AAE 37850/C
ID AAF97850 standard: DNA: 6405 BP.

XX

AC AAF97850;

XX
DT 31-MAY-2001 (first entry)

XX DE Human neuroblastoma cell line NB-1 1p36 nucleotide sequence SEQ ID NO:64

XX Human; chromosome 1; 1p36; neuroblastoma cell line; NB-1; anticancer;
KW tumour suppressor; human 1p36 homozygosity deletion domain; tumour;
KW diagnosis: ds.

XX
XX
-
страницы; на

OS Homo sapiens.

XX
3

PN WO200116311-A1.

XX
XX

PD 08-MAR-2001.

XX

PF 31-AUG-2000; 20

XX

PR 31-AUG-1999;

PR 09-MAY-2000; 20

XX

PA (HISM) HISAMIT

PA (CHIB-) CHIBA P

XX	
FI	Nakagawara A;
XX	
XX	WPI; 2001-226686/23 .
XX	
PT	Human lp36 homozygosity deletion domain from the 36-position of first
PT	chromosome short arm in human neuroblastoma cell lines, applicable e.g.
PT	in gene diagnosis of tumors as well as in developing anti-cancer drugs.
XX	
XX	Example 8; Page 93-95; 226pp; Japanese.
XX	
CC	The present invention describes a homozygosity deletion domain co-
CC	existing in the 36-position of the first chromosome short arm (lp36) in
CC	human neuroblastoma. Also described are base sequences from the lp36
CC	position of human neuroblastoma cell lines (NB-1 and MASS-NB-SCH-1),
CC	which are tumour suppressor genes in human neuroblastoma. The genes are
CC	tumour suppressor genes, base sequence data of which are applicable as
CC	tumour markers and reagents in studying mechanism of tumour body
CC	formation, and gene diagnosis of tumours as well as in developing anti-
CC	cancer drugs. AAF97787 to AAF97829 represent PCR primers used in the
CC	exemplification of the present invention, and AAF97830 to AAF97874
CC	represent sequences given in the exemplification of the present invention
XX	
SQ	Sequence 6405 BP; 1650 A; 1449 C; 1383 G; 1923 T; 0 U; 0 Other;

Query Match	16.7%	Score 416.2;	DB 5;	Length 6405;
Best Local Similarity	59.9%;			
	Pred. No. 7.5e-76;			
Matches 856;	Conservative	1;	Mismatches 504;	Indels 68;
				Gaps 7;

Qy	603	GTGTTTTTTTGTGTTGTTTATAGACAGCGGCTTGCTCTGTCA	CCCGAGGCATGAGCACAG	662
Dd	3636	GTCTTATTTTTTATTTTTTTGAGGTGAGTTTCACTGTTGTGCC <td>CGGTCGAGTGTCAA</td> <td>3577</td>	CGGTCGAGTGTCAA	3577

Qy 663 TGGTGCACCATAGTCACTCAGCCTCAACCTCTGAGCTCAAGGATCTGCTGACCTC 722

Db	3576	TGGCAGATCTCAGCTTACAACAACCTCTGCCTCTGAGTTCAAGTGATTTCTCCTGCTC	3517
Qy	723	AGCCTCCCAAGTACCTGGGACTACGAGCGTGCACACACGCTTGCTTAATTAAAAAAT	782

Db 3516 AGCCTCCGAGTAGCTGGGATTACAGGCATGCACCACTCAGCTGATTTTGTGTA 3457

Qy	783	TTTTTTGTAGAGACTGGGCTTATTACGTTGGCCAGGCTTCTCTTAAACTCCTGGTTCA	842
Db	3456	TTTTTAGTAGAGA-TGGGGTCTCCATGTTGGTCAGGCTAGTCTCAAAC TCCGCACCTCA	3398

[illegible]

DB	3397	GGTGATCCCGCCCTGCGCTCCCGAAGTCTGGGATTACAGGCGTGAGCCCATAGGCC
QY	903	CGGCTACTTATTTCTTTACATTCCATCTTCCATAGAAATGAAGTCCACAGAACAGGG

Db	3337	GGCCATTTTACT	-----	3322
Ov	963	ATTACTGCCTATTTCTTCTCTTTTGGACAGAGTCTCACTTCATCACTCAACC	1022	

[illegible]

Qy	1023	TCGGTTCAGTCACTCACTGCTCCGGGTTCAAGYGATTCCTCGCCTARGCCT	1082
Db <td>3268</td> <td>-----TCGGGTCACTGCAACTCCGGCTCCAGGTTCAAGGATTCCTTTGGCTCAGTCT</td> <td>3214</td>	3268	-----TCGGGTCACTGCAACTCCGGCTCCAGGTTCAAGGATTCCTTTGGCTCAGTCT	3214

Qy 1083 CCTGAGTAGCTGGAATTACAAGCGTGACCAACCATGCTTGGCTAATTTTGTATTTTAA 1142

Db	3213	CCCAAGTAGCTGGGATTACAGGCATGCACCAACCATGCCCCGGCTAA--ATTGTGATTTTTA	3156
Qy	1143	GCAGAGATGGGTTTTACAGATTTGCCCCAGGCTGCTCAAACTCCTGACCTCAAGTGAT	1202

Db 3155 GTAGAGATGGGTTTCTCCATGTTGGTTAGGCTGGTCTCAAACTCCCAACCTC-AGTGAT 3097

QY 1203 CTGCTTCCCTTCAGTCTCCCAAAAGTCTCGAAATTATAGGGCTGAGTCACTGTGCTCTGGCCG
 Db 3096 GCACGAGCCTCGGCTCCCAAAAGTCTGGGATTACAGGCATGAGCACCGCGCCCGGCCG 3037

Qy 1263 ATTACTGCTATTTCTTTTATGCTATATCCCGAGCTCTAGAGCAGTGTCTGCATATAG 1322
Db 3036 TAGGTATCTTTATACAGGATGTTTATACAAATTTCTCTTAGAACAATCTTTAGAAATTA 2977
Qy 1323 TAGGTGCTCAATAAATAATGATGAATGACAGCCTAGATATATAAATCTTTCTTTCTTTT 1382
Db 2976 ATACTAGGAAAAAGGTTT-----CTCTCGAACAATATTAGGCTTTCTTTTACCAGA 2924
Qy 1383 TTTAAAAAATCTTGACAACCTTTGCGAATAAATACAAATCTTGATCTGCTTTTCTACT 1442
Db 2923 GTGAAATTTGACAAATTTATCTTTTCCCTTTTGTGCAATTAATTCATGTTCCAAATTAAT 2864
Qy 1443 TATCACCTGTTATGACTTTTTCATATTTGCTCAACCTTTATGTTACTGTTTTCAT 1502
Db 2863 TACATTAACACATTTGCCAGGCACTATACATTTAGTTTCTTTGTTACATTAATTT 2804
Qy 1503 TGTTACTATTTTACTCTCAATTAATGCTTAATTTGCTTATATACATCTCTCTGCTCCA 1562
Db 2803 TTACTTATTTTATATTTTAAATAGCTTCACTGTATATATTTTGTGTAAGCTGCTCAATCT 2744
Qy 1563 CTTTAGAAGGCCAAATTTACAAATCTGATGAAGCTATGAACCTCTCCCCAGAGAATA 1622
Db 2743 TTCTAGAAGTAGTAACCTGGTATTAATTAATTAATAAATAAATAAATCTCTGTAATGCTAT 2684
Qy 1623 CACACACACACACACTC-ACACACAGTTTTTTTTTTAATGTTTGAACAAACAGAGAA 1681
Db 2683 TCTGCTATTTAACTGTAATATATGTTGTAAGTCAAGTCTTAGTTTCTTAAACAAAT 2624
Qy 1682 ACCTGCATTAAGGATGTTTGTTCATATTAATTAATAAATAAATCACTTGGGCACAGTGAC 1741
Db 2623 AAAACCAACACACACAGAGCTTTTAAAGTTTACATAAAGATCGGCTGGGCATCGTGGC 2564
Qy 1742 TCAAGCTGTAAACACAGTACTTTGGAAGTCCAAAGTGGGTGGATCACTTGAAGTGAGAA 1801
Db 2563 TCACGCTGTAATCCCAACACTTTTGGAGGCCAAGGCGGTGATCACTTGAAGTCAAGGA 2504
Qy 1802 GTTCGAGACGAGCTGTCAATATGTTGAACCTTCTCTACTAAAAATACAAAAATTA 1861
Db 2503 GTTCAAGACCGCTGGCCACATGATGAACGCTGTCTTACTTAAATAACAAAAATTA 2444
Qy 1862 GCTGGGTGTATGATGCTGCTAGTCCAGCTACTCGGAGGCTGAGGCAAGAAAT 1921
Db 2443 GCTGGGTGTGATGGCAATGACTGTAATCCAGCTACTCGGAGGCTGAGGCAAGAAAT 2384
Qy 1922 TGCTTGAACCTGGAGGAGGAGGTTGCAAGTGGCGGAGATCCACACTGCACTCCAGCC 1981
Db 2383 CACTTGAACCCGGAGATGAAGGTTGCAAGTGGCGGAGATCCACACTGCACTCTAGCC 2324
Qy 1982 TGGCGGACACAGCAGACTCTATCTCAAAAAATAAATAAATAAATAA 2030
Db 2323 TGGATGACAGAGCGAGACTCTGCTCAAAAAATAAATAAATAAATAA 2275

RESULT 11
ADP13332/C
ID ADP13332 standard; DNA; 126990 BP.

XX AC ADP13332;

XX DT 26-AUG-2004 (first entry)

XX DE Renal cell carcinoma differentially expressed gene #68.

XX KW ds; diagnosis; non-blood disease; solid tumor; gene expression;
KW peripheral blood mononuclear cell; renal cell carcinoma; prostate cancer;
KW head/neck cancer; differential expression.

XX OS Homo sapiens.

XX FN W02004048933-A2.

XX PD 10-JUN-2004.

XX XX

21-NOV-2003; 2003WO-US037481.
21-NOV-2002; 2002US-0427982P.
03-APR-2003; 2003US-0459782P.
(AMHP) WYETH.
(TWIN/) TWINE N C.
(BURC/) BURCZYNSKI M E.
(TREP/) TREPICCHIO W L.
(DORN/) DORNER A.
(STOV/) STOVER J A.
(SLOW/) SLONI D K.
Twine NC, Burczynski ME, Trepicchio WL, Dorner A, Stover JA;
Sloni DK;
WPI; 2004-460799/43.

Diagnosing non-blood disease such as solid tumor, involves comparing differential expression profile of specific genes in peripheral blood sample of subject with reference expression profile of specific genes.

Disclosure; SEQ ID NO 68; 350pp; English.

The invention relate to a method of diagnosing (M1) non-blood disease such as solid tumor by providing peripheral blood sample of human having non-blood disease, and comparing an expression profile of specific genes in the peripheral blood sample to reference expression profile of the genes, where each of the genes is differentially expressed in peripheral blood mononuclear cells (PBMCs) of patients having the disease as compared to PBMCs of normal humans. The method is useful for diagnosing non-blood disease such as solid tumor. The solid tumor is chosen from renal cell carcinoma (RCC), prostate cancer and head/neck cancer. The peripheral blood sample comprises enriched PBMCs. The peripheral blood sample is a whole blood sample (claimed). (M1) is useful for identifying genes that are differentially expressed in peripheral blood samples isolated at different stages of progression, development or treatment of RCC and/or other solid tumors. This sequence corresponds to a gene that is differentially expressed and detected by the method of the invention. (Note: this sequence is not given as part of the printed specification but was obtained from WIFO in electronic format at ftp.wifo./pub/published_pct_sequences).

Sequence 126990 BP; 36683 A; 27389 C; 27065 G; 35853 T; 0 U; 0 Other;

Query Match 16.8%; Score 416; DB 12; Length 126990;
Best Local Similarity 60.2%; Pred. No. 1.5e-75;
Matches 859; Conservative 1; Mismatches 536; Indels 32; Gaps 9;

Qy 611 TTCTGTTGTTTATAGACAGGGTCTTGTCTGTCCAGGCGATGACACAGTGGTGCAA 670
Db 27332 TTTCTCTAATGCTGAGATGAGTCTACTCTGTCAACCAGGCTGGAGTGCAGTGGTGGA 27273

Qy 671 CCATAGGTCACTGACGCTCAACCTCTGTAGCTCAAGGGATCTGTGACCTCAGCTCCC 730
Db 27272 TCTCA-----CCTCCACCTCCAGGTTCAAGCGATTCTCTGCTCAGCT-CC 27225

Qy 731 AAGTAGCTGGACTACAGGGTGCACACACCGCTGGCTAAATTAATAAATTTTGT 790
Db 27224 TGSTAGCTGGGAATACAGGCGTGCACCAATCCCGGCTAAATTTTATATA--TTTTCAGT 27167

Qy 791 AGAGACTGGGCTTACTTACCTTGGCCAGGCTTCTTAAACTCTGTGCTTCAAGCAATCC 850
Db 27166 AGAGTGCAGTTTCAACAGTGTTCACAGGCTGGTCTCAAACTCTGTGACCTCAAGTATCC 27107

Qy 851 TCCTACTTGGCATCCCAAAAGTCTGGGATTAAGGGGTGAGCCACCATGTGGGGTACT 910
Db 27106 ACCCGCGCAGCTCCCAACGCTGCTGGATAATAGGGGTGAGCCACTGACCTGGCCTTT 27047

Qy 911 TATTTCTTTACATTCATCTTTCCAATAGATTAAGATCCACAGACAGGGATTACTGC 970
Db 27046 CTGTACTCTTTTATTTCTATTTTCACTGGTCTGTTTCTTTTGGTTTCAGTATCATAGTACTTT 26987

Qy 971 CTATTTCTTCCTTCTTTTGTAGACAGAGTCTCACTTCATCACTCAACCTCCGTTCA 1030
Db TGTTCCTGTAGCTTAAATAAT - ATATATTTTAAAGACTGGAATGCAAGTCAATCTCA 26929
Qy 1031 GCTCACTGCAACCTCTGCTCCCGGTTCAAGYATTTCTCTGCTAAGCCTCTCTAGTA 1090
Db TGTTCCTGTAGCTTAAATAAT - ATATATTTTAAAGACTGGAATGCAAGTCAATCTCA 26929
Qy 1091 GCTGGAATTAACAAGCTGCAACCATCTGCTGGCTAATTTTGTATTTTGTAGCAGAT 1150
Db TGTTCCTGTAGCTTAAATAAT - ATATATTTTAAAGACTGGAATGCAAGTCAATCTCA 26929
Qy 1151 GGGGTTTACATGTTCCAGGCTGCTCAAACTCTGACCTCAAGTATCTGCTGC 1210
Db TGTTCCTGTAGCTTAAATAAT - ATATATTTTAAAGACTGGAATGCAAGTCAATCTCA 26929
Qy 1211 CTAGTCTCCCAAGTCTGGAATTAAGGCTGAGTCACTGCTGCTGGCGGATTAATCTGT 1270
Db TGTTCCTGTAGCTTAAATAAT - ATATATTTTAAAGACTGGAATGCAAGTCAATCTCA 26929
Qy 1271 CTATTTTCTTTATTTGCTATATCCAGATCTAGAGCAGTCTGACATATAGTAGTGTCT 1330
Db TGTTCCTGTAGCTTAAATAAT - ATATATTTTAAAGACTGGAATGCAAGTCAATCTCA 26929
Qy 1331 CAATAAATAATGATGAATGACAGCCTAGATATAAACTTTTCTTTTCTTTTAAAC 1390
Db TGTTCCTGTAGCTTAAATAAT - ATATATTTTAAAGACTGGAATGCAAGTCAATCTCA 26929
Qy 1391 AATCTTGACAACTTTGACAGATAAATAAATCAATCTGCTGCTTTTCTCTATCACTT 1450
Db TGTTCCTGTAGCTTAAATAAT - ATATATTTTAAAGACTGGAATGCAAGTCAATCTCA 26929
Qy 1451 TGTATGACCTTTTCAATGCTTCAAACTTTATGTTATCTGTTTCTTTTCTTTTAAAC 1510
Db TGTTCCTGTAGCTTAAATAAT - ATATATTTTAAAGACTGGAATGCAAGTCAATCTCA 26929
Qy 1511 TTTTGTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1570
Db TGTTCCTGTAGCTTAAATAAT - ATATATTTTAAAGACTGGAATGCAAGTCAATCTCA 26929
Qy 1571 GGGCAATTTAACAATCTGATGAAGTATGAACCTCTCCAGAGATAACACACACA 1630
Db TGTTCCTGTAGCTTAAATAAT - ATATATTTTAAAGACTGGAATGCAAGTCAATCTCA 26929
Qy 1631 CACACACACTCAACACAGTCTTTTCTTATGTTTCTTCACTGAGTCAAGCAAGAA 1682
Db TGTTCCTGTAGCTTAAATAAT - ATATATTTTAAAGACTGGAATGCAAGTCAATCTCA 26929
Qy 1683 CTTGCAATTAGAGGATGTTGTTTCAATTAATTAATAAATCACTGAGTGGGCAAGTCACT 1742
Db TGTTCCTGTAGCTTAAATAAT - ATATATTTTAAAGACTGGAATGCAAGTCAATCTCA 26929
Qy 1743 CAAGCTGTAAACACAGTACTTTTGAAGTCCAAAGTGGGTTGATCACTTGGAGTGAAG 1802
Db TGTTCCTGTAGCTTAAATAAT - ATATATTTTAAAGACTGGAATGCAAGTCAATCTCA 26929
Qy 1803 TTCGAGACAGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1862
Db TGTTCCTGTAGCTTAAATAAT - ATATATTTTAAAGACTGGAATGCAAGTCAATCTCA 26929
Qy 1863 CTGGGTTGATGATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1922
Db TGTTCCTGTAGCTTAAATAAT - ATATATTTTAAAGACTGGAATGCAAGTCAATCTCA 26929
Qy 1923 GCTTGAACCTGGAGGAGGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1982
Db TGTTCCTGTAGCTTAAATAAT - ATATATTTTAAAGACTGGAATGCAAGTCAATCTCA 26929
Qy 1983 GGGCGACACAGGAGGATCTATCTCAAAATAAATAAATAAATAAATAAATAAATAAATA 2030
Db TGTTCCTGTAGCTTAAATAAT - ATATATTTTAAAGACTGGAATGCAAGTCAATCTCA 26929

RESULT 12

ACN44170/c
ID ACN44170 standard; DNA; 196686 BP.
XX ACN44170;
AC ACN44170;
XX 18-NOV-2004 (first entry)
DT 18-NOV-2004 (first entry)
XX Human genomic sequence hCG39530.
DE Cytostatic; carcinoma; lymphoma; cancer; human; gene; ss.
XX Homo sapiens.
XX WO2003073826-A2.
XX 12-SEP-2003.
XX 28-FEB-2003; 2003WO-US006235.
XX 01-MAR-2002; 2002US-00087192.
XX (SAGR-) SAGRES DISCOVERY.
XX Morris DW;
XX WPI; 2003-328604/31.
XX Recombinant nucleic acid useful for diagnosis and treatment of carcinoma comprises a nucleotide sequence.
XX Claim 1; SEQ ID NO 484; Opp; English.
XX The present invention relates to novel DNA and protein sequences which are associated with carcinomas. The sequences are useful for: (i) for screening drug candidates; (ii) for screening of bioactive agent capable of binding to Carcino Associated Protein (CAP); (iii) for screening of a bioactive agent capable of modulating the activity of CAP; (iv) for evaluating the effect of a candidate carcinoma drug; (v) for treating carcinoma; (vi) for inhibiting the activity of CAP; (vii) for treating carcinoma; (viii) for neutralizing the effect of CAP; (ix) as a biochip; (x) for diagnosing carcinoma or a propensity to carcinoma; and (xi) for determining Carcino Associated (CA) gene copy number. In addition, the CA genes are useful as DNA vaccines and the CAP are useful as markers of carcinoma including lymphoma. The present sequence is one such CA coding sequence. Note: This patent is an equivalent to basic patent US2002182586A1, for which no sequence data was published
XX SQ Sequence 196686 BP; 53978 A; 42758 C; 43862 G; 55372 T; 0 U; 716 Other;
Query Match 16.6%; Score 414.2; DB 11; Length 196686;
Best Local Similarity 60.7%; Pred. No. 4e-75;
Matches 869; Conservative 1; Mismatches 464; Indels 97; Gaps 8;
Qy 605 TTTTCTTCTGCTGTTTGTAGACAGGCTCTGCTGTCACCCAGGATGAGCAGTG 664
Db TTTTCTTCTGCTGTTTGTAGACAGGCTCTGCTGTCACCCAGGATGAGCAGTG 664
Qy 665 GTGCAACCATAGGTCACTGACGCTCAACCTCTGAGTCTCAAGGGATCTGCTGACCTCAG 724
Db GTGCAATCTCGGCTCACTGCAACCTCCACCTCCAGATTCAGATTCCTCTGCTGCTCAG 98874
Qy 725 CCTCCCAAGTAGCTGGGACTAGAGGTCACCACTGCTGCTGCTGCTGCTGCTGCTGCTGCT 784
Db CCTCCCAAGTAGCTGGGACTAGAGGTCACCACTGCTGCTGCTGCTGCTGCTGCTGCTGCT 784
Qy 785 TTTTGTAGACAGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 844
Db TTTTGTAGACAGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 844
Qy 845 CAATCT 904
Db CAATCT 904
Qy 98755 GGATCCACCCACCTTTGGCTCCCAAGTCTGGGATTCAGGATTCAGGATTCAGGATTCAGG 98696
Db GGATCCACCCACCTTTGGCTCCCAAGTCTGGGATTCAGGATTCAGGATTCAGGATTCAGG 98696

QY	785	TTTTGTAGAGACTGGGTCCTTACTACGTTGGCCAGGCTTGCTTAAACTCCTGGCTTCAAG	844
Db	69050	TTTATTAGACGGGGTTTTGCCATGTTGCCAGGCTGGTCTCAAACTCCTGACCTCAAG	68991
QY	845	CAATCTCTTACCTTTGGCATCCCAAAGTGTCTGGATTTACAGGGGTGAGCCACCATGTGG	904
Db	68990	TGATCCACCGCCCTCGGCTTCCAAAGTGTCTGGATTTACAGGTGGAACACCCGACCTG	68932
QY	905	GCTACTTATTTCTTTACATTCATCTTTTCAATPAGAAATGTAAGATCCACAGAACGGAT	964
Db	68931	GGAGACTTTTATTTTCTTTCTTTCTTTTCTTTTCTTTTCTTTTCTTTTCTTTTCTTTT	68872
QY	965	TACTGCCATATTTTCTTCTTTTCTTTTGTGACAGAGTCTCACTCATCACTCAACCTC	1024
Db	68871	CTCCCTCTGTGCGCCAGGCTGGAGTGGAGTGGCGCAATC-----	68833
QY	1025	CGTTCAAGTCACTGCAACCTCTGCCTCCGGGTTCAAGTGAATTTCTTCCCTTAAGCCTCC	1084
Db	68832	---TCAGCTCACTGCAACCTCCGCTCCGGGTTCAAGCGATTTCTTGCCTCAGCCTCC	68776
QY	1085	TGAGTAGCTGGAAATTACAGGCTGCAACGATGCTTGGCTAAATTTTGTGATTTTATG	1144
Db	68775	CGAGTAGCTGGGACTACAGGCATGTGCCACCAACACCGACTAA-TTTTGTGATTTCTT	68717
QY	1145	AGAGATGGGGTTTTACCATGTGCCCCAGAGCTGTCTCAAACTCTGACTCAAGTGATCT	1204
Db	68716	AGAGACGGGGTTTCAACATATTGGCCAGGCTGTCTCAAACTCTGACTCTGAGGTGATCC	68657
QY	1205	GCCTGCCCTCAGTCTCCAAAGTCTCGGAATATAGGCGTGAAGTCTGCTCTGGCCGAT	1264
Db	68656	ACCCACCTCAGCTCCCAAGTGTCTGGATTTACAGGCTGAACCAACCGGCTGGGCA-	68598
QY	1265	TACTGTCTATTTCTTTATTTGCTATATCCCAAGTCTAGAGCAGTGTCTGACATATAGTA	1324
Db	68597	----CTTTTATTTTCTTAGGCCACTCTGATTAACAAATGACTTGGGTGTGAATGTTGAA	68542
QY	1325	GGTGCTCAATAAATAATTGATGAATGCACGCTAGATATAAATCTTTCTTTTCTTTT	1384
Db	68541	GCACATTCATCCATAGCTTGATCTGTCTCCCTTGGAGCAAGGAAGGGAACTGA-----	68489
QY	1385	TAAACCAACTCTTGACAACCTTTGCAGAAATAAATACAATCTTGCAATCTGCTTTTCACTTA	1444
Db	68488	-GGAACGAACTGGGAAAATGTGCAG-----GGCCACCTTACCTACCGTGTGTGCATC	68436
QY	1445	TCACCTTGTTATGACTTTTTCATATGTCCTCAAACTTTATTTGTTACTGTGTTTTTCATG	1504
Db	68435	CCGAC-AGAAAGGCACTTGTGAAATCGAATACTGGCATTTGTTTCTGTTCTTTTCTG	68377
QY	1505	TTACTATTTTAGTCACTGAATAATATGGCTTAATTTGTTTATACATCCTCTGCTCCACT	1564
Db	68376	----GATCTTGCAGCTGGGTGCGACTTGTGCATACACCACTTGAGTGGAACTGTCGCC	68321
QY	1565	TTAGAAGGCCAAATTTACAAATCTGATGAAGACTATGAACCTCTCTCCCAAGAGAAATACA	1624
Db	68320	GAAAGAAGCCCTGTAGGGTAAATACACAGTGGATTTAAAGACCTGGGCCACTGAGAAACC	68261
QY	1625	CACACACACACACTCACACAGTTTTTTTTTTTAAATGTTTGCACCTAAGACACAGAAACC	1684
Db	68260	AGCGCACTTGCTTTACTACCTTTTTTGTGCTCATTTATTTGTTTCTTTTAAATGAACA	68201
QY	1685	TGCATTTAGAGGATGTTGTTTTCATATTAATTAABATACTCAGTTGGGCAAGTGACTCA	1744
Db	68200	AAATTTAGAGAAAGTTATTC-----CAGGCCAGCATGGTGATCA	68160
QY	1745	AGCCTGTAACCAAGTACTTTGGAACTCCAAGTGGGTGGATCACTTCTGAGGTGAGAAGTT	1804
Db	68159	TGCTCTGATCCAGCACTTTAGGAGCTGAGCGGGTGGATCACTTCTGAGTCTCAGGATT	68100
QY	1805	CGAGACAGGCTGGTCAATATGGTGAACCCCTATCTTACTATAAATACAAAAATTTAGCT	1864
Db	68099	CGAGACAGGCTAGCCAAATGGTGAACCCCTGCTCTACTAAAAATACAAAAATTTAGCC	68040

QY	1865	GGGTAGTATGTCATGCCTGTAGTCCCGAGCTACTCGGGAGGCTGAGGCAAGAATTGC	1924
Db	68039	GGGCATGGTGATGTTAAGTCTGTAGTCCCGAGCTACTCAGGAGGCTGAGTAGGACAAATCAC	67980
QY	1925	TTGAACCTGGGAGGCGAGAGGTTGCAGTGAAGCGAGATCCCACCACATGCACCTCCAGGCTGG	1984
Db	67979	TTGAACCCMGAGGCGGAGGTTGCAGTGAAGCGAGATTTGCGCATGCATCTCCAGGCTGC	67920
QY	1985	GGCACACGCGAGACTCTATCTCAAAAAATAAATAAAAAATAAAGGATCGGAGAGAA	2044
Db	67919	GGCAGAGTGAAGCTCCATCTCAAAAAATAAATAAAAAATAAAGGAGAAAGAA	67860
 RESULT 14 ADQ97523/c			
ID	ADQ97523	standard; DNA; 215974 BP.	
XX	AC	ADQ97523;	
XX	DT	07-OCT-2004 (first entry)	
XX	DE	Human cancer associated sequence HD09-008, SEQ ID 500.	
XX	KW	Cytostatic; Gene Therapy; cancer; leukemia; lymphoma; Human; ds.	
OS	Homo sapiens.		
PN	WO2004060304-A2.		
PD	22-JUL-2004.		
XX	PP	22-DEC-2003; 2003WO-US041389.	
XX	PR	27-DEC-2002; 2002US-00330773.	
XX	PA	(SAGR-) SAGES DISCOVERY INC.	
XX	PI	Morris DW, Malandro MS;	
XX	DR	WPI; 2004-543781/52.	
XX	PT	New isolated cancer associated nucleic acids comprising at least 10	
PT	contiguous nucleotides, useful for diagnosing, preventing and/or treating		
PT	cancers such as leukemia and lymphoma.		
PS	Claim 1; SEQ ID NO 500; 199pp; English.		
XX	The present invention relates to cancer associated sequences (ADQ97025-		
CC	ADQ98004). The sequences are useful for the diagnosis, prevention and/or		
CC	treatment of cancer, such as leukemia and lymphoma. Note: The sequence		
CC	data for this patent did not form part of the printed specification, but		
CC	was obtained in electronic format directly from WIPO at		
CC	ftp.wipo.int/pub/published_pct_sequences.		
XX	Sequence 215974 BP; 56601 A; 42897 C; 46516 G; 65586 T; 0 U; 4374 Other;		
SQ	Query Match	16.5%; Score 413.6; DB 12; Length 215974;	
	Best Local Similarity	59.6%; Pred. No. 5.4e-75;	
	Matches	847; Conservative 0; Mismatches 529; Indels 46; Gaps 7;	
QY	610	TTTGTGTGTTTGTAGACAGGGTCTTGCTGTGTCAACCAGGATGAGCAGTGTGTGCA	669
Db	4846	TGTCCTTTTTTTTTGAGACAGGGTCTGCTCTGTTGCCAGGCTGAGTACAGTGGCACA	4787
QY	670	ACCATAGTCTACTGAGGCTCAACTCTCTGAGCTCAAGGATCTGCTGACCTCAGCTCC	729
Db	4786	ATCATGGCTCATTTGAGGCTCAAACCTCTGGGCTCAAAATGAGCCTCCAACCTCAGCTCC	4727
QY	730	CAAGTAGCTGGAGCTPACGAGGCTGCACCACGAGCTGCTTAATTAATAAAAATTTTTTTG	789
Db	4726	TAAGTAGCTGGAATACAGGCAATACATCACCACTCTAGCTAATTTTTGTATTTTT---	4670
QY	790	TAGAGACTGGGTCTTACTACGTTGGCCAGGCTTGCTTAAACTCTCTGGCTTCAAGCAATC	849

[illegible]

Db	3632	TGTTGGCGCGTCTGTAATCCAGCTACTCGGGAGGCTGAGACAGGAGAAATTCCTTGAA	3573
Qy	1930	CCTGGGAGGCGAGGTTGCAGTGCAGCCGAGATCCACCACTGCATCCAGCCTGGGCGAC	1989
Db	3572	CCCAGGAGCGGAGGTTGCAGTGCAGCAGAGATTGCACCTGTTGCACCTCCAGCCTGGGAGAC	3513
Qy	1990	ACAGCGAGACTCTATCTCAAAAAAAAAATAAATAAATAA	2031
Db	3512	AGAGCAAGACTCCGCTCTCAAAAAAGAAACAGAGGAGAAACAGA	3471
RESULT 15			
AAL42369/c			
XX	ID	AAL42369 standard; DNA; 13670 BP.	
XX	AC	AAL42369;	
XX	XX		
XX	28-JUN-2002	(first entry)	
XX	Human	Guanine nucleotide binding protein gamma 7 (GNG7) gene sequence.	
XX	Human; gene; ds;	guanine nucleotide binding protein gamma 7; G protein;	
XX	GNG7; novel polymorphic site; drug screening; gene therapy;		
XX	GNG7-related disease; pancreatic cancer; GNG7 haplotyping;		
XX	GNG7 genotyping.		
XX	Homo sapiens.		
XX	OS	Homo sapiens.	
XX	Key	Location/Qualifiers	
FT	allele	replace(3900, T)	
FT		/*tag= a	
FT		/note= "Single nucleotide polymorphism"	
FT	allele	replace(3975, G)	
FT		/*tag= b	
FT		/note= "Single nucleotide polymorphism"	
FT	CDS	4005..9671	
FT		/*tag= c	
FT		/product= "Guanine nucleotide binding protein gamma 7"	
FT	exon	4005..4085	
FT		/*tag= d	
FT		/number= 1	
FT	intron	4086..9545	
FT		/*tag= e	
FT		/number= 1	
FT	exon	9546..9671	
FT		/*tag= f	
FT		/number= 2	
FT	allele	replace(9958, A)	
FT		/*tag= g	
FT		/note= "Single nucleotide polymorphism"	
XX	WO200218647-A1.		
XX	PN		
XX	PD		
XX	07-MAR-2002.		
XX	23-AUG-2001;	2001WO-US026279.	
XX	25-AUG-2000;	2000US-0228234P.	
XX	PA	(GENA-) GENAISSANCE PHARM INC.	
XX	Finkel K, Kliem SE, Koshy B;		
XX	WPI; 2002-315548/35.		
XX	P-P5DB; AAO14753.		
XX	Novel genetic variants of guanine nucleotide binding protein, gamma 7		
XX	gene useful in studying expression and function of the protein, and for		
XX	screening drugs to treat diseases e.g. pancreatic cancer.		
XX	Claim 1; Fig 1; 62pp; English.		
XX	The invention comprises the nucleotide and amino acid sequences of the		

human guanine nucleotide binding protein (G protein) gamma 7 (GN07). The invention specifically relates to the discovery of three novel polymorphic sites in the GN07 gene. The GN07 nucleotide and protein sequences are useful for screening for drugs which target GN07 and may be used to treat (gene therapy) GN07-related diseases (e.g. pancreatic cancer). The GN07 nucleotide sequence can be used for haplotyping GN07, genotyping GN07 and predicting the haplotype pair for GN07 in an individual. The present DNA sequence represents the human GN07 gene sequence.

SQ Sequence 13670 BP; 3047 A; 3946 C; 3680 G; 2997 T; 0 U; 0 Other;

Query Match 16.5%; Score 413.4; DB 6; Length 13670;

Best Local Similarity 60.6%; PRED. NO: 3.3e-75;
Matches 876; Conservative 1; Mismatches 512; Indels 56; Gaps 10;

Qy	504	TTTTTTTTTGTGTTGTTT	AGACACAGGGTCTT	GTCTCTGTCTACCCAGGCAAT	-GAGCACAG	662
Db	5900	TTTGTGTTTTTGTGTTTTTCT <td>AGATGAGAGTCT <td>CACCCCTGT <td>CGTCTGGGTTGGAGGGCAG</td> <td>5841</td> </td></td>	AGATGAGAGTCT <td>CACCCCTGT <td>CGTCTGGGTTGGAGGGCAG</td> <td>5841</td> </td>	CACCCCTGT <td>CGTCTGGGTTGGAGGGCAG</td> <td>5841</td>	CGTCTGGGTTGGAGGGCAG	5841

Qy	663	TGTTGCAACCAATAGGTCACTGCAGCCTCAACCTCTGAGCTCAAGGATCTGTGACCTC	722
Db <td>5840</td> <td>TGTCGCAATCTCAGCTCTGCTGAACTCTGCTCTCGCGGTCAGAGCTATTCCTCTGATC</td> <td>5781</td>	5840	TGTCGCAATCTCAGCTCTGCTGAACTCTGCTCTCGCGGTCAGAGCTATTCCTCTGATC	5781

QY 723 AGCCTCCCAAGTAGCTGGGACTACGAGCGTGCAACACCGCTGGCTAATTAATAAAT 782

QY 783 TTTTCTCTAGAGACTGGGTCTTACTACGTTGGCCAGGCTGTCTTAACCTCTGGCTTCA 842

843 ACCAATCCCTCTACCTTGGGCATCCCAAGTGTCTGGGAATTA CAGGGGTGAGCCACCATGTG 902

Qy CGGCTACTTATTTCTTTACATTCCCATCTTTCCAATAGAATGTCAAGATCCACAGAACAGGG 962

DB	3602	1GGCCAGATATAAGTTTTTTTATAGAAGTTCCTGTGAAGGTGTGTTC	5555
QY	963	ATTACTGCCTATTTTCTTCTTTTCTTTTGGAGACAGAGTCTCACTTCATCACTCAACC	1022

	DB	:	5553	-TGTATTCTTTTTTTTTTTTTTTTGATAGAGTCTCGCTCTGTTCCGCCAGCT	5495
	QY	:	1023	TCCGT-----TCAGTCACTGCAACCTCTGCTCCC CGGTTTCAAGYGATTC	1068

Accession	Gene	Accession	Gene
D8	5494	QY	1069

Db	5434	TCCTGCCTCAGCCTCCCAAAATGCTGGATTACAGGCAATGACCAATCCGGGTAAAT	5375
Qy	1128	TTTCTTTTGTAATTTTACGACAGATGGGGTTTTTACATGTTTCCGAGGGCTGCTCAAA	1184

Accession	Gene	Accession	Gene
5374	Db	1185	Qy
CTTTTCTTTGTAATTTTAGGAGAGACGGGGTTTCACCATGTTGGCCAGGCTGGTCTCGAA		CTCCTGACCTCAAGTGA TCTGCTGCTGCTCAGTCTCCCAAAGFGCTGGAATTTAGGCGTG	

Db	5314	CCCCGACCTCAGGTGATCTGCCCACTCAGCTTCCCAAGTGTCTGGAAATATAAGTGTG	5255
Qy	1245	AGTCAGTGTGCTGGCCGATTA	1304

Db	5254	AGTCACGCGCTCAGCCTGTT-TTTTATTTTGTAGATGCGGTCTTGTCTATGTTGTGC	5196
Ov	1305	GCAGTGCTGCATATAGTAGTGCTCAATAAATAATTGATCAATGCACAGCCTAGATAT	1364

[illegible]

Db
5135 TGGGATT-----ACAGCTGTGAGCCCACTGTGCTCACATGCAAAATTTT 5092

Db	5091	TTTTTTTTTTTGATGTGATTTCACATT-----TAAATCCGTAGACTTTCAGACACAGCAGGT	5036
Qy	1485	TTGTTACTGTTTTTTCATTGTTTACTATTTTAGTCACTGAATAATATGGCTTAAATTTGCTT	1544
Db	5035	GACCGTCTGTAATCCAGATGGCCCTCATCTAATCAGTTGAAGGCTTAAGAGAAAGCC	4976
Qy	1545	ATACATCCTCTGCTCACTTTTAGAGGCCAAATTTACAAATCTGATGAAGATTAAGAAC	1604
Db	4975	TGGGGTCTTCCAGGAGGAGGAACTCTGCTCCAGGTGGCCCTTGGGGTTCACAGCTGCAG	4916
Qy	1605	CCTCTCCCCAGAGAAATACACACACACACACACTCACACACAGTTTTTTTTTAAATGTT	1664
Db	4915	CACCAACTCTTCCTGGGTCTCCAGCCTGCTGGCCTACACTGCAGATTTTCAGAGCTTGC	4856
Qy	1665	TGCNACTTAACACAAGAAACCTGCATTAGAGGATGTTTGTTTCATATTAAATTTAAATAA	1724
Db	4855	TTAAATTAATACTTCGCAA--GGCGCAGCCTCAAGTGTGATCCTGTGTAAAAAATCTC	4797
Qy	1725	CAGTTGGGCACAGTACTCAAGCCTCTTAACCAACAGTACTTTTGGAAAGTCCAAGGTGGGTG	1784
Db	4796	GGGCCAGGCATGTTGGCTCAGCCTTGGATCCAGGACTTTGGAGGCCGAGGACGCCG	4737
Qy	1785	ATCATCTGAGGTGAGAAGTTTCGAGACAGCCTGGTCAATATGTTGTAACCCCTATCTCTAC	1844
Db	4736	ATCACCTGAAGTTAGGAGTTTCAAGACAGCCTGGGCAACATGGCGAAAGCCGCTCTCTAC	4677
Qy	1845	TAAAAATAC--AAAAATTTAGCTGGGTCTGAGTGATGCATGCTGTAGTCCAGCTACTTCGG	1903
Db	4676	TAAAAATACAAAAAATTAGCCAGGCATGGTGGTGCATGCTGTATATCCAGCTATTGGG	4617
Qy	1904	AGGCTGAGGCAAGAGAAATGCTTTGAACCTGGGAGGCAGAGGTTGCAGTCAGCCGAGATCC	1963
Db	4616	AGGTTGAGGCAGAGAATTTGCTTTGAACCCGGGAGGCAGAGGTTGCAGTCAGCCGAGATCG	4557
Qy	1964	CACCACTGCACCTCCAGCCTGGGGGCACACAGCAGACTCTATCTCAAAAAAATAAATAAAT	2023
Db	4556	CGCACTGCACTCCAGCCTGGGTGACAGAGCAAGACTCCGTCTCAAAACAACAAACAATA	4497
Qy	2024	AAAAAT 2028	
Db	4496	CAAAAT 4492	

Search completed: February 9, 2006, 18:54:29
Job time : 1619.05 secs

THIS PAGE BLANK (USPTO)

GenCore version 5.1.7
Copyright (c) 1993 - 2006 Bioceleration Ltd.

OM nucleic - nucleic search, using sw model

Run on: February 9, 2006, 15:02:28 ; Search time 9346.29 Seconds
(without alignments)
12519.885 Million cell updates/sec

Title: US-10-607-806-1-G7328-T9182_COPY_7000_9500
Perfect score: 2499.4
Sequence: 1 gtcgtgtcactgtgtccag.....tttgagaccagcctggacaa 2501

Scoring table: IDENTITY NUC
Gapop 10.0 , Gapext 1.0

Searched: 41078325 seqs, 2339354128 residues

Total number of hits satisfying chosen parameters: 82156650

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : EST.*

- 1: gb_est1.*
- 2: gb_est2.*
- 3: gb_est3.*
- 4: gb_est4.*
- 5: gb_est5.*
- 6: gb_est6.*
- 7: gb_est7.*
- 8: gb_est8.*
- 9: gb_est9.*
- 10: gb_est10.*
- 11: gb_est11.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	505.4	20.2	507	5	BU785040 in43g09.x
2	344.2	13.8	4087	4	BC024593 Homo sapi
3	341	13.6	3990	4	AL713681 Homo sapi
4	337.2	13.5	15970	9	AQ839852 260L13-C5
5	315.8	12.6	897	6	CD557847 AGENCOURT
6	311	12.4	3095	4	CR858326 Pongo pyg
7	307.2	12.3	5797	4	CF749233 Homo sapi
8	301.2	12.1	1605	4	CR860521
9	301.2	12.1	1863	4	CR599842
10	300.6	12.0	946	5	BQ958903
11	298	11.9	918	5	BQ706343
12	293.6	11.7	736	8	CX780936
13	290.6	11.6	3092	4	CR613629
14	288.2	11.5	855	5	BQ681302
15	287.6	11.5	769	5	BQ710315
16	285.2	11.4	4088	4	HSM808459
17	283	11.3	3165	4	CR859576 Pongo pyg
18	281.6	11.3	800	1	AU120942
19	281.4	11.3	3259	4	CR860168
20	280.4	11.2	672	6	CA431692
21	280	11.2	3474	4	HSM800886
22	278.8	11.2	658	9	AQ393450 CITBI-E1-

c 23	277.6	11.1	877	9	AQ739838
c 24	277.2	11.1	1875	4	BC009270
c 25	276.5	11.1	3990	4	HSM803026
c 26	276.4	11.1	4828	4	HSM802759
c 27	276.2	11.1	881	5	BU521286
c 28	274.8	11.0	3552	4	CR860263
c 29	273.8	11.0	736	6	CA427039
c 30	273.4	10.9	629	5	BX509360
c 31	273.4	10.9	2230	4	CR859082
c 32	272.4	10.9	652	6	CA427045
c 33	272.4	10.9	666	6	CA431783
c 34	272.4	10.9	815	8	CX785622
c 35	272.2	10.9	922	5	BU501973
c 36	272	10.9	617	3	BI861844
c 37	271.8	10.9	644	7	CN480313
c 38	271.6	10.9	1042	5	EX377759
c 39	271.6	10.9	2097	4	CR614786
c 40	271.4	10.9	603	1	AL707313
c 41	271.2	10.9	5325	4	HSM804049
c 42	271.2	10.9	5785	4	HSM802309
c 43	270.6	10.8	2330	4	CR615928
c 44	270.4	10.8	618	9	BZ609884
c 45	270	10.8	685	5	BU633001

ALIGNMENTS

RESULT 1
BU785040/c
LOCUS in43g09.x1 HR85 islet Homo sapiens cDNA clone IMAGE:6125008 3',
DEFINITION mRNA sequence.
ACCESSION BU785040
VERSION BU785040.1 GI:23830576
KEYWORDS EST.
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini; Homnidae; Homo.
REFERENCE 1 (bases 1 to 507)
AUTHORS Melton,D., Brown,J., Kenty,G., Permutt,A., Lee,C., Kaestner,K., Lemishka,I., Searce,M., Brestelli,J., Gradwohl,G., Clifton,S., Hillier,L., Marra,M., Pape,D., Wylie,T., Martin,J., Blistain,A., Schmitt,A., Theising,B., Ritter,E., Ronko,I., Bennett,J., Cardenas,M., Gibbons,M., McCann,R., Cole,R., Tsagarishvili,R., Williams,T., Jackson,Y. and Bowers,Y.
TITLE Endocrine Pancreas Consortium
JOURNAL Unpublished (2000)
COMMENT Contact: Douglas Melton, Klaus H. Kaestner, & Hiroshi Inoue
Endocrine Pancreas Consortium
Harvard University, Howard Hughes Medical Institute
Dept of Molecular and Cellular Biology, 7 Divinity Ave, Cambridge, MA 02138
Tel: 617-495-1812
Fax: 617-495-8557
Email: dmelton@bioh.harvard.edu
Library was constructed by Dr. Hiroshi Inoue DNA sequencing by: Washington University Genome Sequencing Center For information on obtaining a clone please contact: Dr. Hiroshi Inoue (hinoue@im.wustl.edu)
Seq primer: -40UP from Gibco
High quality sequence stop: 443.
Location/Qualifiers
1. 507
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/clone="IMAGE:6125008"
/issue_type="Purified pancreatic islet"
/lab_host="DH10B"
/clone_lib="HR85 islet"

/note="Organ: Pancreas; Vector: pBluescript SK(-); Site_1: NotI; Site_2: XhoI; cDNA made by oligo-dT priming. Size-selected on agarose gel. Average insert size ~1kb. 5' XhoI site was destroyed after directional cloning. Amplified once. Contact information: Hiroshi Inoue, MD, Metabolism Div. (Alan Permutt Lab), Washington University School of Medicine, Box 8127, 660 South Euclid Ave., St. Louis, MO 63110, E-mail: hinoue@imgate.wustl.edu, Tel: 314-362-1916, Fax: 314-747-2692."

ORIGIN

Query Match 20.2%; Score 505.4; DB 5; Length 507;
 Best Local Similarity 99.8%; Pred. No. 2.4e-35;
 Matches 506; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

1506 TACTATTTTGTCTCAATTAATATGCTTAATTTGCTTATACATCTCTCTGCTCCACTT 1565
 |||||
 Db 507 TACTATTTTGTCTCAATTAATATGCTTAATTTGCTTATACATCTCTCTGCTCCACTT 448
 |||||

1566 TAGNAGGCCAAATTTACAAATCTGATGAAGCTATGAACCTCTCTCCAGAGAAATACAC 1625
 |||||
 Db 447 TAGNAGGCCAAATTTACAAATCTGATGAAGCTATGAACCTCTCTCCAGAGAAATACAC 388
 |||||

1626 ACACACACACACTCACACACAGTCTTTTATTTTATGTTTGCACCTAAAGACAAGAAACCT 1685
 |||||
 Db 387 ACACACACACACTCACACACAGTCTTTTATTTTATGTTTGCACCTAAAGACAAGAAACCT 328
 |||||

1686 GCATTAGAGGATGTTGTTTCATATTAATTAATAATAAATACAGTTGGGCACAGTACTCAA 1745
 |||||
 Db 327 GCATTAGAGGATGTTGTTTCATATTAATTAATAATAAATACAGTTGGGCACAGTACTCAA 268
 |||||

1746 GCCTGTACACACAGTACTTGGAGTCCAGGTGGGTGGATCACTTCAGGTGAGAGTTC 1805
 |||||
 Db 267 GCCTGTACACACAGTACTTGGAGTCCAGGTGGGTGGATCACTTCAGGTGAGAGTTC 208
 |||||

1806 GAGACACAGCTGTCATATGTTGAAACCCCTATCTCTACTATAAATAACAAAATTTAGCTG 1865
 |||||
 Db 207 GAGACACAGCTGTCATATGTTGAAACCCCTATCTCTACTATAAATAACAAAATTTAGCTG 148
 |||||

1866 GGTGTAGTATGATGCTGTAGTCCAGCTACTTCGGAGGCTGAGGCAAGAGAAATGCT 1925
 |||||
 Db 147 GGTGTAGTATGATGCTGTAGTCCAGCTACTTCGGAGGCTGAGGCAAGAGAAATGCT 88
 |||||

1926 TGAACCTGGAGGCGAGGTTGCAGTGGCGAGATCCACCACTGCACCTCCAGCTGGG 1985
 |||||
 Db 87 TGAACCTGGAGGCGAGGTTGCAGTGGCGAGATCCACCACTGCACCTCCAGCTGGG 28
 |||||

1986 CGACACAGCGAGACTCTATCTCAAAA 2012
 |||||
 Db 27 CGACACAGCGAGACTCTATCTCAAAA 1
 |||||

RESULT 2

BC024593
 LOCUS BC024593 4087 bp mRNA linear HTC 29-JUN-2004
 DEFINITION Homo sapiens cDNA clone IMAGE:3914314, with apparent retained intron.

ACCESSION BC024593

VERSION 1 GI:22137609

KEYWORDS HTC.

SOURCE Homo sapiens (human)

ORGANISM

Homo sapiens
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini;
 Hominidae; Homo.

REFERENCE 1 (bases 1 to 4087)

Strausberg, R.L., Feingold, E.A., Grouse, L.H., Derge, J.G.,
 Klausner, R.D., Collins, F.S., Wagner, L., Shenmen, C.M., Schuler, G.D.,
 Altschul, S.F., Zeeberg, B.S., Buetow, K.H., Schaefer, C.F., Bhat, N.K.,
 Hopkins, R.F., Jordan, H., Moore, T., Max, S.I., Wang, J., Heien, F.,
 Klapchenko, L., Marusina, K., Farmer, A.A., Rubin, G.M., Hong, L.,
 Stapleton, M., Soares, M.B., Bonaldo, M.F., Casavant, T.L.,
 Scheetz, T.E., Brownstein, M.J., Usdin, T.B., Toshiyuki, S.,

Carninci, P., Prange, C., Raha, S.S., Loquellano, N.A., Peters, G.J.,
 Abramson, R.D., Mullaly, J., Bosak, S.A., McEwan, P.J.,
 McKernan, K.J., Malek, J.A., Gunaratne, P.H., Richards, S.,
 Worley, K.C., Hale, S., Garcia, A.M., Gay, L.J., Hulyk, S.W.,
 Villalon, D.K., Munz, D.M., Sodergren, E.J., Lu, X., Gibbs, R.A.,
 Fahey, J., Helton, E., Kettman, M., Madan, A., Rodriguez, S.,
 Sanchez, A., Whiting, M., Madan, A., Young, A.C., Shevchenko, Y.,
 Bouffard, G.C., Blakesley, R.W., Touchman, J.W., Green, E.D.,
 Dickson, M.C., Rodriguez, A.C., Grimwood, J., Schmutz, J., Myers, R.M.,
 Butterfield, Y.S., Krzywinski, M.I., Skalska, U., Smalios, D.E.,
 Schnerch, A., Schein, J.E., Jones, S.J. and Marra, M.A.

Generation and initial analysis of more than 15,000 full-length
 human and mouse cDNA sequences

Proc. Natl. Acad. Sci. U.S.A. 99 (26), 16899-16903 (2002)

12477932

2 (bases 1 to 4087)

Strausberg, R.

Direct Submission

Submitted (01-MAR-2002) National Institutes of Health, Mammalian
 Gene Collection (MGC), Cancer Genomics Office, National Cancer
 Institute, 31 Center Drive, Room 11A03, Bethesda, MD 20892-2590,
 USA

NIH-MGC Project URL: <http://mgc.nci.nih.gov>

Contact: MGC help desk

Email: cgabs-r@mail.nih.gov

Tissue Procurement: ATCC

cDNA Library Preparation: Life Technologies, Inc.

DNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)

DNA Sequencing by: Sequencing Group at the Stanford Human Genome
 Center, Stanford University School of Medicine, Stanford, CA 94305

Web site: <http://www.shgc.stanford.edu>

Contact: (Dickson, Mark) mcd@paxil.stanford.edu

Dickson, M., Schmutz, J., Grimwood, J., Rodriguez, A., and Myers,

R. M.

Clone distribution: MGC clone distribution information can be found
 through the I.M.A.G.E. Consortium/LLNL at: <http://image.llnl.gov>
 Series: IRAK Plate: 22 Row: m Column: 14
 This clone was selected for full length sequencing because it
 passed the following selection criteria: Hexamer frequency ORF
 analysis

This clone has the following problem: retained intron.

FEATURES

source

1..4087
 /organism="Homo sapiens"
 /mol_type="mRNA"
 /db_xref="taxon:9606"
 /clone="IMAGE:3914314"
 /tissue_type="uterus, leiomyosarcoma"
 /clone_lib="NIH_MGC_71"
 /lab_host="DH10B"
 /note="Vector: pCMV-SPORT6"

ORIGIN

Query Match 13.8%; Score 344.2; DB 4; Length 4087;
 Best Local Similarity 58.3%; Pred. No. 6.8e-22;
 Matches 859; Conservative 1; Mismatches 544; Indels 69; Gaps 12;

599 TTGGGTTTTTTTTTTGTTGTTTGTAGACAGAGGTTCTGTGTACCCAGGATGAGC 658
 |||||
 Db 2019 TTTGMAATTTGTTTGTGTTTGTGATGACAGGTTCTGGGTTGCCACCCAGGCTGGAGT 2078
 |||||

659 ACAGTGGTGAACATAGTCTCACTGAGCTCAACCTCTGAGCTCAAGGATCTGCTGA 718
 |||||
 Db 2079 GTAGTGGTGTGATCATAGTCTCAATGAGCCCCAACCTTCGGGGCGCAAGTGATCTCA 2138
 |||||

719 CCTCAGCTCCCCAAGTAGCTGGGACTACGAGCGTGCACACACCGCTGGCTTAATATAA 778
 |||||
 Db 2139 CCTCAGCTCTCAGTAGCTGGGACTACAGCATGTGCCACACACACCTCTCTAATATA 2198
 |||||

779 AAAT-----TTTTTTGTAGAGACTGGGTCTTACTACGTTGGCCAGGCTGTGCTTAAACTC 833
 |||||
 Db 2199 TTTTATTTTATTTTGTAGAGATGAGGTTTTCATTATGTTGCCAGAGCTGGTCTGGAACTC 2258
 |||||

ORIGIN		YRCBEGKAFKHSKLTTHKR FHTGKPYRCE	
Query Match		12.3%; Score 307.2; DB 4; Length 5797;	
Best Local Similarity		56.9%; Pred. No. 9e-19;	
Matches 902; Conservative		1; Mismatches 569; Indels 114; Gaps 14;	
Qy	605	TTTTTTTGTGTTTGTAGACAGGCTTCTGCTCTGTCACCCAGCGATGACAGTG	664
Dy	4976	TTTTTTTCTTTTTTTTGTGACAGAGTCTGCTCTGTTGCCAGAGCTGGAGTGCA	4917
Qy	665	GTGCAACCATAGTCACTCGAGCGCTCAACCTCTGAGCTCAAGGGAATCTGCTGACCTCAG	724
Dy	4916	GCACGATCTCGGCTCACTGCAACCTCTGCTCTCTGGGTACACCAATCTCCG-CTCAG	4958
Qy	725	CCTCCCAAGTAGCTGGGACTACGAGCGTGACACCAAGCGCTGGCTAAATTAATAAATTT	784
Dy	4857	CCTCTGGAGTAGCTGGGACTACAGCGCCACCAACCGCCAGCTAAATTTA-----TT	4804
Qy	785	TTTTGTAGAGACTGGGTCTTACTACGTTGGCCAGGCTTGCTTAAACTCTCTGGCTTCAAG	844
Dy	4803	TTTAGTAGACAGAGGTTTCAACGTGTAGCCAGATGGTCTTGATCTCTCTGACCTCATG	4744
Qy	845	CAATCTCTCTACCTTGGCATCCCAAGTGTGGGATACAGGGTGAGCCACCATGTGCG	904
Dy	4743	--ATCCACTCGCTGGGCTCCCAAGTGTGGGATACAGGTGTGAGCCACTGTCCCTG	4686
Qy	905	GCTACTTAATTTCTTCAATTCATCTCTTCCAAATGAATGAAGATCCACAGAACAGGGAT	964
Dy	4685	GCITTTTTCTTCTT-----TTTGAGATGGAGTGTCCCTCTGT	4648
Qy	965	TACTGCCTATTTCTCTCTCTTTTGTGACAGAGTCTCACTTCATCACTCAACCTC	1024
Dy	4647	CACCCAGCTGGAGTGAGTGGTCTAT-----	4620
Qy	1025	CGTTTCACTCACTCAACCTCTGCTCCCGGTTCAAGYGATCTCTGCTCAAGCTCC	1084
Dy	4619	--CTCAGCAGACTGAACCTCACTCCAGGTTCAAGCCATCTCTGCTCGGCTTCC	4562
Qy	1085	TGAGTAGCTGGAATTAAGCGGTGACACCATGCTTGGCTGCTAAATTTTGTATTTTACG	1144
Dy	4561	CGAGTAGCTAGGAATAC--GGGTGACACCATGCCCAACTAAATTTTGTGTTTTT	4508
Qy	1145	AGAGATGGGTTTACCATGTTGCCAGGCTGCTCAACTCTGACCTCAAGTGATCT	1204
Dy	4507	-AGTAGAGATTTTCAATATGTTGTGAGGCTGGTCTTGAACCTCTGACCTCA--TGATCT	4451
Qy	1205	GCCTGCTCTCACTCCCAAGTGTGGAATATATAGGCGTGAGTCACTGTGCTGGCGGAT	1264
Dy	4450	GCCGACGTCACTACCAAGTGTGGATTTACAGGATGAGCCACTGGTCCAGCCCA--	4392
Qy	1265	TACTGTCTATTTCTTTATGCTATATCCCGAGATCTAGAGCAGTGTCTGACATATAGTA	1324
Dy	4391	-----TATTCCTATTTATTTCTGTATAATTTTATTTATGACCAATAAAATTCACCTGTA	4337
Qy	1325	GGTCTCAATAAATGATGATGACACCTCTAGATATAAATTTCTTTCTTTCTTTT	1384
Dy	4336	GTCAATTAACAATTAATTTGGACATTTAAAG--TAACATAAAGTGATATAATACACTGTT	4279
Qy	1385	TAAACATCTTGACAACTTTTGACAGATAAATACAATCTGCTATCTGCTTTTTCACCTTA	1444
Dy	4278	TGGAATGAAGAGGATNAATGCTAGAGGTGATAGATACCTTATTTACCTAATGATATAC	4219
Qy	1445	TCACCTGTTATGACTTTTTCATATG-----CCTCAAACTTTATTTGTTACTGTTTTT	1499
Dy	4218	TACATATTTGATGCTGAATCAAAATATTTCTTATAAGACATATATACATATATATA	4159
Qy	1500	CATTGTTACTATTTTACTGCTGATATATATGCTT---AATTTGCTTATACATCTCTCT	1556
Dy	4158	CTCAAAATAGTAATAATAAATTTAGAAAAAAGATTTAAATTTAACTAACGGGAACAAT	4099
Qy	1557	GCTCACCTTTAGAAAGGCCAAATTTTACAAATCTGATGAAGCTATGAACCTCTCTCCCGAGA	1616

Db	4098	ATTCTTTTTCAGTTTTCAGTTTACTGGCAAGTGATTTACTAGTGATGACATTCACACTACATACC	4039
Qy	1617	GAAATACACACACACACACTCACACAGTTTTCATGTTTTCATG-----	1662
Dy	4038	AAATAGTACTTCTTCCATCTTTTCTTACACCAATTTTGAGTAAGCCCGATAGGTAA	3979
Qy	1663	-----TTTGGCACTAAGACAAGAACCTTCATTTAGAGGATGTTTCTTCATATTAATA	1716
Dy	3978	GTTAGTGGCGTATAATGCTTCATTAATGCAATAGTTTAAAGATGCAAAAATATA	3919
Qy	1717	AAATAACTCAGTTGGGCACAGTGAATCAAGCCTGTAAACACAGTACTTTGGAGTCCAAG	1776
Dy	3918	ATAAATTTAGGCTGGACATATTTGGCTCAACCTTAATCCCAACACTTTTGAGAGCCGAG	3859
Qy	1777	GTGGTGGATCACTTCAGGTGAGAGTTGAGACCTGCTGATCATATGTTGAAACCTT	1836
Dy	3858	GTGGTGGATCGCTGAGCTCAGGATTTGAGACTAGCTGACCAATATGGGGAACCCC	3799
Qy	1837	ATCTCTACTAAATAAATAAATAAATTTAGCTGGTGTAGTGATGCTGTAGTCCCAGCT	1896
Dy	3798	ATGCTATTAATAAATAAATAAATTTAGCCAGGCTGTGTACGCACTGTAGTCCCACCT	3739
Qy	1897	ACTCGGAGGCTGAGCAAGAGATTTGCTTGAACCTGGGAGGAGGTTGCAAG-----TG	1952
Dy	3738	ACTCGGAGGCTGAGGAGAGATTTCTTGAACCTGGGAGGTTGAGGTTGCAAGTCTTG	3679
Qy	1953	AGCGGATCCCACTGCACTCCAGCTGGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG	2012
Dy	3678	TGCTGAGACTGCACTCTCTCCCTCCAGCTGGGTGACAGGCAAGACTCCATCCCCAAAG	3619
Qy	2013	ATAAATAAATAAATAAAGGATCGGAGAGAAACAAATAAATAAGATCTCTCAAGGATA	2072
Dy	3618	AAAAGTTAGTAATAAATAAATAAATTAAGTTTACACATAATCTACAATTTTAAATAT	3559
Qy	2073	GCAGAGATAGTAAATATATGTAATAAAGTTTAAATGCAATTTAACTGTAATCTTATG	2132
Dy	3558	ATTGCAATTTTATTCAGTAAAGTAAATTTAGTAAATAAATTTGATATATCTTAAATTA	3499
Qy	2133	TTTATTTTGGTTATAAAGTAAACAA 2158	
Dy	3498	TTTCTTTCTCTCAGTATAATGTAGAA 3473	
RESULT 8			
CR860521/c			
LOCUS			
DEFINITION			
Pongo pygmaeus mRNA; cdna DKFZp468P2419 (from clone DKFZp468P2419).			
ACCESSION			
CR860521			
VERSION			
CR860521.1 GI:55731876			
KEYWORDS			
HTC.			
SOURCE			
Pongo pygmaeus (orangutan)			
ORGANISM			
Pongo pygmaeus			
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;			
Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini;			
Hominoidea; Pongo.			
1 (bases 1 to 1605)			
REFERENCE			
AUTHORS			
Ortenwaelder,B.; Obermaier,B.; Deutschenbaur,S.; Schapp,A.,			
Mewes,H.W., Weil,B., Amid,C., Osanger,A., Fobo,G., Han,M. and			
Wiemann,S.			
The German CDNA Consortium			
Direct Submission			
Submitted (12-NOV-2004) MIPS, Ingolstaedter Landstr.1, D-85764			
Neuherberg GERMANY			
COMMENT			
Clone from S. Wiemann, Molecular Genome Analysis, German Cancer			
Research Center (DKFZ); Email s.wiemann@dkfz-heidelberg.de;			
sequenced by Medigenomix (Martinsried/Germany) within the CDNA			
sequencing consortium of the German Genome Project.			
This clone (DKFZp468P2419) is available at the RZPD Deutsches			
Ressourcenzentrum fuer Genomforschung GmbH in Berlin, Germany.			
Please contact RZPD for ordering:			
http://www.rzpd.de/cgi-bin/products/cl.cgi?CloneID=DKFZp468P2419			
Further information about the clone and the sequencing project is			
available at http://mips.gsf.de/projects/cdna/.			

Qy 1159 ACCATGTTGCCAGGCTGCTCAAACTCTGACCTCAAGTGTGCTGCTGCTCAGTCT 1218
 Db |||||
 Qy 1036 ACTGTGTAGGTAGGTGGTCTCGATCTCTGACCTCA--TGATCCATCGCCTTGGCCT 1093
 Db |||||
 Qy 1219 CCCTAAAGTGTGGAATATATAGGCGTGAAGTCACTGTGCTGCGCCGATTAATCTATTTTC 1278
 Db |||||
 Qy 1094 CTCAAGTGTGGATTTACAGGCATGAGCACCGCAGCTGGCTG-----TTTT 1141
 Db |||||
 Qy 1279 TTTATGCTATATATCCAGATCTAGACAGTGTCTGACATATATAGTGTGCTCAATAAT 1338
 Db |||||
 Qy 1142 TTTATTTCTATATGATTTGTTCTGTGGAATGTCACTGAGTGTAGCGGTGTTTTTCAATG 1201
 Db |||||
 Qy 1339 AATTGATGATGACAGCGCTAGATATAAATCTTCTTTCTTTTAAACATCTTGA 1398
 Db |||||
 Qy 1202 TTTATAAAGCTGAATGTTCTGTATGTCACTGACGACCCCTAGTGTCACTGCTACT 1261
 Db |||||
 Qy 1399 CAACCTTTGCGAATAAATAACAATCTGCTATCTGCTTTTTCATCTATCACTTGTATGA 1458
 Db |||||
 Qy 1262 CTCTGCTGTGATAGATGGGGCTGTCCACCCCAATCTCTGGGTTCCGCTGCTC 1321
 Db |||||
 Qy 1459 CTTTTTCATATGCTCAAACTTTATTTGTTACTGTTTCTTTTCAATGTTACTATTTAGTC 1518
 Db |||||
 Qy 1322 GCACCTGAAGGTACCCACAGATGTGTCTGCTGCTCCATCCACCTTCTTGTGT 1381
 Db |||||
 Qy 1519 ACTGAATATATGCTTTAATTTGCTTATATATCTCTGCTGCTCCATTTAGAGCCCAAT 1578
 Db |||||
 Qy 1382 TCTCATGTTCTGCTC-----CCTAGAGACCAATTTCTCTCTACAGCAGTAG----- 1428
 Db |||||
 Qy 1579 TTACAAATCTGATGAAGCTATGAACCTCTCCCGAGAGAAATACACACACACACACA 1638
 Db |||||
 Qy 1429 -----TTTCAATAGAAATACATGCGAGACATA 1459
 Db |||||
 Qy 1639 CTCACACACAGTTTTTTTAAATGTTTGCATTAAGACAAACCTGCTATAGAGATG 1698
 Db |||||
 Qy 1460 C-----GTCAATTTTAGATTTTCTAGTAGTCACTTTAGAAAAGTGAAAAGAGGCCA 1510
 Db |||||
 Qy 1699 TTTGTTTATATTAATAATAACTCAGTTGGGACAGTCACTCAAGCTGTAAACACA 1758
 Db |||||
 Qy 1511 GCTG-----CAGTGGCTCAACCTGTAAATCCCA 1538
 Db |||||
 Qy 1759 GTACTTTGGAAGTCCAAAGTGGTGGATCACTTGAGGTGAGAAGTTCCGAGACCAAGCTGG 1818
 Db |||||
 Qy 1539 GCATTTTGAAGGCTGAGTGGTGGATCAC--GAGTCCAGGAGATCGAGACCATCTGG 1596
 Db |||||
 Qy 1819 TCAATATGGTGAACCCCTATCTCTACTATAAATAAC--AAAAATTAGCTGGGTGATGATG 1877
 Db |||||
 Qy 1597 CCAACATGGTGAACCCCTCTCTACTATAAATAACAAAAATTAGCTGGGCGTGGTGGCA 1656
 Db |||||
 Qy 1878 CATGCTGTAGTCCAGCTACTCGGAGGCTGAGGCAAGAGATTTGCTTGAACCTGGAG 1937
 Db |||||
 Qy 1657 GCGCCTGTAGTCCAGCTACTCGAGAGGCTGAGGCAAGAGATTTGCGGTGAACCCGTAG 1716
 Db |||||
 Qy 1938 GCAGAGGTTGAGTGGAGGAGTCCACCACTGCTCCAGCTGCGGCGACACAGCGAG 1997
 Db |||||
 Qy 1717 GCAGAGCTTGAGTGGAGGAGATAGGCGCACTGCTCCAGCTTGGCGGAGAGAGAG 1776
 Db |||||
 Qy 1998 ACTCTATCTCAAAAAATAATAATAATAAATGAAGGATCGGAGAGA 2043
 Db |||||
 Qy 1777 ACTCCATCTCAAAAAATAATAATAATAAATGAAGGATCGGAGAGA 1822
 Db |||||

RESULT 10
 BQ958903
 LOCUS BQ958903 10035485 NIH_MGC_40 Homo sapiens cdna clone IMAGE:6483214
 DEFINITION 5', mRNA sequence.
 ACCESSION BQ958903
 VERSION BQ958903.1 GI:22374381
 KEYWORDS EST.
 SOURCE Homo sapiens (human)
 ORGANISM Homo sapiens
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini;
 Hominidae; Homo.
 1 (bases 1 to 946)
 NIH-MGC http://mgi.nci.nih.gov/.
 National Institutes of Health, Mammalian Gene Collection (MGC)
 Unpublished (1999)
 Contact: Robert Strausberg, Ph.D.
 Email: cgabbs@mail.nih.gov
 Tissue Procurement: DCTB/DRP
 cdna Library Preparation: Rubin Laboratory
 DNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
 DNA Sequencing by: Agencourt Bioscience Corporation
 Clone distribution: MGC clone distribution information can be
 found through the I.M.A.G.E. Consortium/LLNL at:
 http://image.llnl.gov
 Plate: LLCM2667 row: g column: 23
 High quality sequence stop: 534.
 Location/Qualifiers
 1..946
 /organism="Homo sapiens"
 /mol_type="mRNA"
 /db_xref="taxon:9606"
 /clone="IMAGE:6483214"
 /tissue_type="carcinoma, cell line"
 /lab_host="DH10B (phage-resistant)"
 /clone_lib="NIH MGC 40"
 /notes="Organ: prostate; Vector: pOTB7; Site: 1: XhoI;
 Site 2: EcoRI; cdna made by oligo-dt priming.
 Directionally cloned into EcoRI/XhoI sites using the
 following 5' adaptor: GGCACGAG(G). Library constructed by
 Ling Hong in the laboratory of Gerald M. Rubin (University
 of California, Berkeley) using ZAP-cDNA synthesis kit
 (Stratagene) and Superscript II RT (Life Technologies).
 Note: this is a NIH_MGC Library."

ORIGIN

Query Match 12.0%; Score 300.6; DB 5; Length 946;
 Best Local Similarity 69.1%; Pred. No. 1.1e-17;
 Matches 468; Conservative 1; Mismatches 197; Indels 11; Gaps 4;
 Qy 596 AGTTTGGGTTTTTTTTTTGTTGTTTGTAGACAGAGGTTTCTGTCTGTCACCCAGGATG 655
 Db |||||
 Qy 24 AGNTTTTTTTTTTTTTTTTTTTTGTAGACAGAGTCTGTCTGTCTCACCAGGCTGG 83
 Db |||||
 Qy 656 AGCACAGTGGTGAACCATAGTCACTGACGCTCAACCTCCTGAGCTCAAGGAGTCTGC 715
 Db |||||
 Qy 84 AGTGCAGTGGTG---CGATCGCTCACTGCAACCTCAGCTCTCTGGATTTAAGTGTCTC 140
 Db |||||
 Qy 716 TGACCTCAGCTCCCAAGTAGCTGGAGCTACGAGCTGACACCAACCAAGCTGCTAATTA 775
 Db |||||
 Qy 141 CTGCTCAGCTCCCAAGTAGCTGGATTTACAGTGTGACCAACCATGCCGCTAATTT 200
 Db |||||
 Qy 776 AAAAAATTTTTTGTAGAGCTGGGTCTTACTAGTGTGGCCAGGCTTGTCTTAACTCCT 835
 Db |||||
 Qy 201 TTTTGTATTTTGTAGTGAGACAGGGTTTCCATATGTTGGCCAGGCTGCTCGAATCTCT 260
 Db |||||
 Qy 836 GGCTTCAAGCATCTCTACCTTGTGATCCCAAGTGTGGGATTTACAGGGGTGAGCCA 895
 Db |||||
 Qy 261 GACCTC--GTGATCCGCTGCTCCAGCTCCCAAGTGTGGGATTTACAGGTGTGAGCCA 318
 Db |||||
 Qy 896 CCAATGTGGGCTACTTATTTCTTTTCAATTCATCTTTTCCAAATAGAATGTAAAGATCCACAG 955
 Db |||||
 Qy 319 CCAGCTGGCCCCAGTCCAAATATTTAAAGATTTTCTTCTTAGTGTCTTGAAGTTTG 378
 Db |||||
 Qy 956 AACAGGAGTATCTGCTATTTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCT 1015
 Db |||||
 Qy 379 CACA-----AAATCTTTTTTTTGTAGATGAGTCTCACTCTGTCACTCCAGGCTGGAGTGC 433
 Db |||||
 Qy 1016 CTCAACTCGTTTCACTGCTCACTGCTGCTCCCGGTTTCAAGYGATTTCTTCTTCTTCT 1075
 Db |||||
 Qy 434 AGTGGGCTGATCTTGGCTCACTGCAACCTCTGCTCTGCTGGTTCAGCAATTTCTCCACC 493
 Db |||||
 Qy 1076 TAAGCTCTCTGAGTAGTGGAATTAACAAGCGTGCACCAACCATGCTTGGCTTAATTTTGT 1135
 Db |||||

Db 494 TCAGCCTCCCAAGTAGCTGGGATATACAGACGTGTGCCACCATACCTGGGTAA-TTTTTCG 552
|||||
Qy 1136 ATTTTATGACAGATGGGTTTACCATGTGCCAGGCTGGTCTCAAACTCCAGCTC 1195
|||||
Db 553 ATTTTATGAGAGGAGGTTTACCATGTGGCAGGGTGGTCTTGAACCTCCAGCTC 612
|||||
Qy 1196 AAGTGATCTGCTCCCTCAGTCTCCCAAGTGTGGAAATTATAGGCGTGAAGTCACTGTGC 1255
|||||
Db 613 AGGTGATCTCTCGCTCGGCTCCCAAGTGTGGGATTACAGGCATGAGCCACCGTGC 672
|||||
Qy 1256 CTGCGCCATTAAGTCTCT 1272
|||||
Db 673 TCAGCGCCANAAATTCCT 689
|||||

RESULT 11
BQ706343/c

LOCUS BQ706343 918 bp mRNA linear EST 16-JUL-2002
DEFINITION AGENCOURT 8484983 NIH_MGC_113 Homo sapiens cDNA clone IMAGE:6301091
5', mRNA sequence.

ACCESSION

BQ706343

VERSION

BQ706343.1 GI:21845242

KEYWORDS

EST.

SOURCE

Homo sapiens (human)

ORGANISM

Homo sapiens

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini;

Hominiidae; Homo.

1 (bases 1 to 918)

NIH-MGC <http://mgs.nci.nih.gov/>.

National Institutes of Health, Mammalian Gene Collection (MGC)

Unpublished (1999)

Contact: Robert Strausberg, Ph.D.

Email: cgapbs-r@mail.nih.gov

Tissue Procurement: Dr. Mark Watson

cDNA Library Preparation: Rubin Laboratory

DNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)

Clone distribution by: Agencourt Bioscience Corporation

Clone distribution: MGC clone distribution information can be

found through the I.M.A.G.E. Consortium/LLNL at:

<http://image.llnl.gov>

Plate: LCM2516 row: g column: 12

High quality sequence stop: 632.

FEATURES

source

1..918

/organism="Homo sapiens"

/mol_type="mRNA"

/db_xref="taxon:9606"

/clone="IMAGE:6301091"

/lab_host="DH10B (phage-resistant)"

/clone_lib="NIH_MGC_113"

/notes="Organ: spleen; Vector: pORF7; Site: 1: XhoI; Site 2:

EcoRI; cDNA made by oligo-dT priming. Directionally cloned

into EcoRI/XhoI sites using the following 5' adaptor:

GGCAGCAG(G). Library constructed by Ling Hong in the

Laboratory of Gerald M. Rubin (University of California,

Berkeley) using ZAP-cDNA synthesis kit (Stratagene) and

Superscript II RT (Life Technologies). Note: this is a

NIH_MGC Library."

ORIGIN

Query Match 11.9%; Score 298; DB 5; Length 918;

Best Local Similarity 68.6%; Pred. No. 1.8e-17;

Matches 456; Conservative 1; Mismatches 197; Indels 11; Gaps 3;

Qy 598 TTTGGGTTTTTTTTTGTGTGTTTATGAGACAGGGTCTGTCTGTCCACCGCATGAG 657

|||||

Db 666 TTTTGGATTATGATTATTTCTTTGAAGCCAGGCTCGCTCTGTCACNACGCGCGAG 607

|||||

Qy 658 CACAGTGGTCAACCATAGTCTCACTGAGCCTCAACTCTCTGAGTCAAGGGATCTGCTG 717

|||||

Db 606 TGCAGTGGTCAATCATGCTCACTCCAGCCTTTATCTCTCTGGGCCCAAGTGATCTCTCA 547

Qy 718 ACCTCAGCCTCCCAAGTAGCTGGGACTACGAGCGTGCACCAACGCGCTGGCTA----- 771
|||||
Db 546 ACCTCAGCCTCCCAAGTAGCTGGGANTATAAGTGGCTACGATGTCTGGCTAAATTTT 487
|||||
Qy 772 -ATTAAAAAAATTTTTTTTGTAGAGACTGGGTCTTACTACGTGGCCAGGCTGTCTTTAAA 830
|||||
Db 486 TATTTATAAATATTTTTTTGTAGAGACAGGCTCTCGCTTTGTGTCGCCAGGCTGGTCTCAA 427
|||||
Qy 831 CTCTGGCTTCAAGCAATCTCTCTACCTTGGGATCCAAAGTCTGGGATACAGGGGTG 890
|||||
Db 426 TTCTTGGACTCAGGCGATCTCTCTGCTTCAAGCTCCCAAAGTCTGAGATTACAGGGGTG 367
|||||
Qy 891 AGCC---ACCATGTGCGGCTACTTATTTCTTTACATTTCCATCTTTTCCAAATGAAATGTAAG 947
|||||
Db 366 AACCTTTCAACTTTTAAATGCTTTTCAATATTTTCCATCTCTGATGAAATTTGGGTTT 307
|||||
Qy 948 ATCCACAGAACAGGGATTAATGCTTATTTTCTTCTTCTTTTGGAGACAGAGTCTCAC 1007
|||||
Db 306 TAAATTTATTTTGTGTTGTTTCTTTTGTGAGACAGAGTCTTGTCTGTCAACCGGC 247
|||||
Qy 1008 TTCAATCACTCAACCTCCGTTTCACTTCACTTCACTTCACTTCACTTCACTTCACTTCACT 1067
|||||
Db 246 TAGAATGCAAGGGGTGTGATCTCGGCTCACTGCAACCTCTGCTCTCGCTTCAAGCGATT 187
|||||
Qy 1068 CTCTCTCCCTAAAGCTCCTCTGAGTAGCTGGAATTACAAAGCTGACACCACTGCTTGGCTAA 1127
|||||
Db 186 CTGTGCTCTCAGCTCTCTCTGAGTAGCTGGATTTACAGCGCTGCTGATGTCCTCAACTAA 127
|||||
Qy 1128 TTTTGTGTTATTTTGTAGCAGAGATGGGTTTACCATTGTTGCCAGGCTGTCTCAAACTC 1187
|||||
Db 126 TTTTGTGTTATTTTGTAGTAGAGACAGTGTTCACCATGTTGGCCAGGCTGTCTCAAACTC 67
|||||
Qy 1188 CTGACCTCAAGTATCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGAG 1246
|||||
Db 66 CTGACCTCAAGTATCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGAG 7
|||||
Qy 1247 TCACT 1251
|||||
Db 6 CCACT 2

RESULT 12

CX780936

LOCUS

CX780936

DEFINITION

HES3_2.B04.b1.A036 NIH_MGC_260 Homo sapiens cDNA clone

IMAGE:30928759 3', mRNA sequence.

ACCESSION

CX780936

VERSION

CX780936.1

GI:58297726

KEYWORDS

EST.

SOURCE

Homo sapiens (human)

Homo sapiens

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini;

Hominiidae; Homo.

1 (bases 1 to 736)

NIH-MGC <http://mgs.nci.nih.gov/>.

National Institutes of Health, Mammalian Gene Collection (MGC)

Unpublished (1999)

Other ESTs: HES3_2.B04.g1.A036

Contact: Daniela S. Gerhard, Ph.D.

Office of Cancer Genomics

National Cancer Institute / NIH

Bldg. 31 Rm10A07 Bethesda, MD 20892

Email: cgapbs-r@mail.nih.gov

Tissue Procurement: BresaGen, Inc.

cDNA Library Preparation: Express Genomics, Inc.

DNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)

DNA Sequencing by: Laboratory for Genomics and Bioinformatics,

University of Georgia

Clone distribution: MGC clone distribution information can be

found through the I.M.A.G.E. Consortium/LLNL at:

<http://image.llnl.gov>

Plate: NDAM1172 row: c column: 08
Seq primer: M13-21 (TGTAAACGACGCCGAGT)
High quality sequence stop: 711
POLYA=Yes.

FEATURES		Location/Qualifiers
source	1..736	
	/organism="Homo sapiens"	
	/mol_type="mRNA"	
	/db_xref="taxon:9606"	
	/clone="IMAGE:30928759"	
	/sex="male"	
	/tissue_type="human embryonic stem cells"	
	/cell_type="human embryonic stem cells"	
	/cell_line="BG01"	
	/lab_host="DH10B-T1 phage-resistant E. coli"	
	/clone_lib="NIH MGC 260"	
	/notes="Vector: pExpress-1; Site.1: NotI; Site.2: EcoRV; RNA obtained from human embryonic stem cells isolated from the inner cell mass of blastocyst stage embryos. Cell line id and NIH Registry designation is BG01. Positive for SSBs, SSBsA4, Tra 1-60, Tra 1-81, CD9, Alk Phos, Oct4 and Nanog expression; negative for SSEA1 expression. Passage number 21. cDNA primed using oligo-dT primer: 5'-pGACTAGTCTAGATCGGAGCGGCCCT(T)25-3' and cloned into the EcoRV/NotI sites of pExpress-1. This primary library is non-normalized (normalized primary library is NIH MGC 261). It was constructed by Express Genomics (Frederick, MD). Sequence ends have been trimmed to exclude vector and regions below Phred quality 16. Note: this is a Mammalian Gene Collection library."	

ORIGIN

Query Match	11.7%; Score 293.6; DB 8; Length 736;
Best Local Similarity	66.9%; Pred. No. 5.1e-17;
Matches 464; Conservative	1; Mismatches 220; Indels 9; Gaps 3;
Qy 599	TTGGGTTTTTTTTTGTGTTGTTTGTAGAGACAGGGTCTTGCTCTGTCAACCAGGATGAGC 658
Db 1	TTTTTTTTTTTTTTTTTTTTTTTTTTGAGATGGAGTCTCACTCTTGTCCAGGCTGAAT 60
Qy 659	ACAGTGTGCAACATAGGTCACTGAGCCTCAACCTCTGAGCTCAAGGGATCTGCTGA 718
Db 61	GCAGTGGCGGGATCTCGGGTCACTGCAACCTCTGCCTCTGGGTTCAAGCACATCTTCTG 120
Qy 719	CCTCAGCCTCCCAAGTAGCTGGGACTACGAGCGTGACACACCGCTCGCTGAATTAATA 778
Db 121	CTTCAGCCCCCTGAGTAGCTGGGATACAGCACCCACACAGGCGCTGATACTT--TT 178
Qy 779	AAATTTTTTTGTAGAGACTGGGCTTACTACGTTGGCCAGGCTTGCTTAAACTCTGTGC 838
Db 179	TTGTATTTTTAGTAAAAATGGGTTTGCTATATTGGCCAGGCTGGTCTCAAACTCCTGAC 238
Qy 839	TTCAAGCAATCTCTACCTTGGCATCCCAAGTGTGGGATACAGGGGTGAGCCACCA 898
Db 239	CTCAGGTGATCCACCTCGCTCCTCAAGCTGCCAAAGTGTGGGATACAGGCATGAACCA 298
Qy 899	TGTGCGGCTACTTATTTCTTTTACATTCATCTTTTCCAATAGATGTAAAGATCCACAGA 958
Db 299	TGCCAGCCTCTTATTTCTTTTAAATATATACAAAGTATTTATTTATTTATTTATTTA 358
Qy 959	AGGGATTACTGCGCTATTTCTTCTCTTTCTTTTGTAGACAGGTCTCACTTCATCACCTC 1018
Db 359	TTTTATTTATTTTGTAGATGGAGTTTTCCTTGTTCGCCAAGCTGGAGTGAATGSCAC 418
Qy 1019	AACCTCCGTTCACTGCTCATCTGCACTCTGCTCCCGGGTTCAAGYGATTTCTCTGCGCTAA 1078
Db 419	GATC-----TCAGCTCACTGCAACTCCACCTCCCGGGTTCAAGTGATTTCTCTGCGCTCA 473
Qy 1079	GCCTCCTGAGTACGTGGAATTACAAGCGGTGCACCACCATGCTTGGCTAAATTTTTTGAT 1138
Db 474	GCCTCCTGAGTAGCTGGGATTCAGGGCACATGCCATCATGCTCGGCTTAA--TTTTGAT 531
Qy 1139	TTTAGCAGAGATGGGGTTTTTACCATGTTTGCCAGGCTGGTCTCAAACTCTCGACCTCAAG 1198

REFERENCE
 1 (bases 1 to 769)
 NIH-MGC http://mgi.nci.nih.gov/
 National Institutes of Health, Mammalian Gene Collection (MGC)
 Unpublished (1999)
 JOURNAL
 COMMENT
 Contact: Robert Strausberg, Ph.D.
 Email: cga@bbs-r@mail.nih.gov
 Tissue Procurement: Dr. Mark Watson
 CDNA Library Preparation: Rubin Laboratory
 CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LNL)
 DNA Sequencing by: Agencourt Bioscience Corporation
 Clone Distribution: MGC clone distribution information can be
 found through the I.M.A.G.E. Consortium/LNL at:
 http://image.lnl.gov
 Plate: LICM2476 row: d column: 21
 High quality sequence stop: 549.
 Location/Qualifiers
 1. 769
 /organism="Homo sapiens"
 /mol_type="mRNA"
 /db_xref="taxon:9606"
 /clone="IMAGE:6282212"
 /lab_host="DH10B (phage-resistant)"
 /clone_lib="NIH MGC 113"
 /note="Organ: spleen; Vector: pOTB7; Site 1: XhoI; Site 2:
 EcoRI; cDNA made by oligo-dT priming. Directionally cloned
 into EcoRI/XhoI sites using the following 5' adaptor:
 GGACGAG(G). Library constructed by Ling Hong in the
 laboratory of Gerald M. Rubin (University of California,
 Berkeley) using ZAP-cDNA synthesis kit (Stratagene) and
 Superscript II RT (Life Technologies). Note: this is a
 NIH_MGC Library."

ORIGIN
 Query Match 11.5%; Score 287.6; DB 5; Length 769;
 Best Local Similarity 67.0%; Pred. No. 1.6e-16;
 Matches 453; Conservative 1; Mismatches 215; Indels 7; Gaps 3;
 QY 604 TTTTGTGAGAGCTGGGCTTCTACGTTGGCCAGGCTTGTCTTAACTCTGGCTTCAA 843
 DB 768 TTTTGTGAGAGCTGGGCTTCTACGTTGGCCAGGCTTGTCTTAACTCTGGCTTCAA 532
 QY 664 GGTGCAACATAGCTGAGCTCAACCTCTGAGCTCAAGGGATCTGCTGACCTCA 723
 DB 708 GGGGTGATCTCGGCTCACTGCAACTCTGCTCTCTGGTTCAAGGATCTCTGCTGCTCA 649
 QY 724 GCCTCCCAAGTAGTGGGACTACGAGCTGACCCACCGCTGGCTTAATTAATAAATT 783
 DB 648 GCCTCCCGAGTAGTGGGACTATAGGGCCCGGCACTGTGCGGCTAATTTTATA--T 592
 QY 784 TTTTGTGAGAGCTGGGCTTCTACGTTGGCCAGGCTTGTCTTAACTCTGGCTTCAA 843
 DB 591 TTTTGTGAGAGCTGGGCTTCTACGTTGGCCAGGCTTGTCTTAACTCTGGCTTCAA 532
 QY 844 GCAATCTCTTACCTTGGCTCCCAAGTCTGGGATTAAGGGGTGAGCCACCATGTGC 903
 DB 531 GTGATCCGCTTGGCTTGGCTTGGCTTGGCTTGGCTTGGCTTGGCTTGGCTTGGCT 472
 QY 904 GGCTACTTATTTCTTTACATTTCTTCCAAATAGAAATGTAAGATCCACAGAACAGGA 963
 DB 471 GGCTGTGTTACATTTTGTGGTCTCTACAGGCTCTGAGGCTCTGCTCATTTTGTGT 412
 QY 964 TTACTGCTATTTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTT 1023
 DB 411 TTTTGTGAGAGCTGGGCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTT 352
 QY 1024 CCGTTTACGCTCACTGCAACCTCTGCTCCCGGTTTCAAGYATTCTCTGCTTAAGCTC 1083
 DB 351 AATCTGGCTCACTGCAACCTCTGCTCCAGATTCAGGATTCCTTGGCTCAGCTC 292
 QY 1084 CTGAGTAGCTGGAAATTAACAGCGTGCACCATGCTTGGCTAA--TTTTTGTATTTT 1141
 DB 291 CCAAGTAGCTGGGATTAACAGCGATGCACCATGCTTGGCTAA--TTTTTGTATTTT 232

QY 1142 AGCAGAGATGGGTTTTACCATGTTGCCAGGCTGTCTCAAACTCTGACCTCAAGTGA 1201
 DB 231 AGTAGAGATGGGTTTTGACCATGTTGCCAGGATGGTCTAGATCTCTTACCTC--GTGA 174
 QY 1202 TCTGCTGCTCAGTCTCCCAAAGTCTCGAAATTTATAGCGTGAATCACTGTGCTGGCC 1261
 DB 173 TCCAACCACTTGGGCTTCCCAAAGTCTGAGATTTACAGGCTGAGCACTGCGCTGGCC 114
 QY 1262 GATTACTGCTATTTT 1277
 DB 113 GGCTCTCTCTTCTTAT 98

Search completed: February 10, 2006, 00:58:23
 Job time : 9357.29 secs

GenCore version 5.1.7
Copyright (c) 1993 - 2006 Bioceleration Ltd.

OM nucleic - nucleic search, using sw model

Run on: February 9, 2006, 14:43:53 ; Search time 12458.7 Seconds
(without alignments)
11410.963 Million cell updates/sec

Title: US-10-607-806-1-G7328-T9182_COPY_7000_9500

Perfect score: 2499.4

Sequence: 1 gtcgtgcaactgtgtccag.....tttgagaccagcctggacaa 2501

Scoring table: IDENTITY NUC
Gapop 10.0 , Gapext 1.0

Searched: 5883141 seqs, 28421725653 residues

Total number of hits satisfying chosen parameters: 117626282

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

- GenEmbl.*
1: gb_ba.*
2: gb_in.*
3: gb_env.*
4: gb_on.*
5: gb_ov.*
6: gb_pat.*
7: gb_ph.*
8: gb_pr.*
9: gb_ro.*
10: gb_sts.*
11: gb_sy.*
12: gb_un.*
13: gb_vi.*
14: gb_htg.*
15: gb_pl.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	2499.4	100.0	8368	8 AY438977	AY438977 Homo sapi
2	2499.4	100.0	122302	8 AC003982	AC003982 Homo sapi
3	2499.4	100.0	220384	14 AC078926	AC078926 Homo sapi
4	2499	100.0	13612	6 AX377239	AX377239 Sequence
5	2110	84.4	189729	14 AC073930	AC073930 Homo sapi
6	462.6	18.5	178887	14 AC068981	AC068981 Homo sapi
7	462.6	18.5	190162	8 AC009244	AC009244 Homo sapi
8	460.6	18.4	123192	14 AP000589	AP000589 Homo sapi
9	460.4	18.4	150266	8 AC022408	AC022408 Homo sapi
10	459.4	18.4	167996	8 AC021753	AC021753 Homo sapi
11	457.4	18.3	137693	8 AC003689	AC003689 Homo sapi
12	456.2	18.3	201460	8 AP003721	AP003721 Homo sapi
13	451.2	18.1	195986	14 AC116933	AC116933 Papio anu
14	449.2	18.0	134210	8 AC005052	AC005052 Homo sapi
15	447.8	17.9	131943	8 AC005484	AC005484 Homo sapi
16	447.8	17.9	136119	8 HS1028D15	AL121886 Human DNA
17	447.8	17.9	186197	14 AL627384	AL627384 Homo sapi
18	446.2	17.9	163681	8 AL136992	AL136992 Human DNA

19	445	17.8	117336	8 AC093171	AC093171 Homo sapi
20	442.6	17.7	91927	8 AC004771	AC004771 Homo sapi
21	441.8	17.7	200430	8 AC011500	AC011500 Homo sapi
22	439.4	17.6	180076	14 AC022177	AC022177 Homo sapi
23	438.8	17.6	201227	14 AC160565	AC160565 Pan trogl
24	438.6	17.5	163704	14 AC141415	AC141415 Pan trogl
25	438.2	17.5	199045	14 AC149081	AC149081 Pan trogl
26	437.8	17.5	185574	14 AC161822	AC161822 Pan trogl
27	435.8	17.4	203790	8 AC010422	AC010422 Homo sapi
28	435	17.4	119043	8 AC139451	AC139451 Homo sapi
29	435	17.4	163542	14 AC129071	AC129071 Pan trogl
30	434.8	17.4	96995	8 AC008149	AC008149 Homo sapi
31	434.2	17.4	152659	8 AL591503	AL591503 Human DNA
32	433.8	17.4	199384	8 AP003531	AP003531 Homo sapi
33	433.2	17.3	89448	8 AC005067	AC005067 Homo sapi
34	433.2	17.3	178976	14 AL391095	AL391095 Homo sapi
35	432.6	17.3	156933	8 AC020750	AC020750 Homo sapi
36	432.2	17.3	81117	8 AC016898	AC016898 Homo sapi
37	432.2	17.3	192961	8 AP002770	AP002770 Homo sapi
38	432	17.3	153392	8 AL133548	AL133548 Human DNA
39	432	17.3	235183	14 AC160951	AC160951 Colobus g
40	431.4	17.3	214031	14 AC160952	AL049759 Human DNA
41	431	17.2	134957	8 HSJ930L11	AC004517 Homo sapi
42	430.8	17.2	44062	8 AC004517	AC004517 Homo sapi
43	430.8	17.2	118396	8 AC073138	AC073138 Homo sapi
44	430.8	17.2	162701	14 AC073317	AC073317 Homo sapi
45	430.6	17.2	41613	8 AC004790	AC004790 Homo sapi

ALIGNMENTS

RESULT 1	AY438977	Homo sapiens phospholipase A2, group IB (pancreas)	8368 bp	DNA	linear	PRI 29-OCT-2003
LOCUS	AY438977	complete cds.				
DEFINITION	AY438977.1	GI:37953284				
ACCESSION	AY438977					
VERSION						
KEYWORDS						
SOURCE						
ORGANISM						
REFERENCE						
AUTHORS						
TITLE						
JOURNAL						
COMMENT						
FEATURES						
source						
repeat_region						
variation						
variation						
repeat_region						
misc_feature						

```
gene /note="Region not scanned for variation"
875. .6548
/gene="PLA2G1B"
mRNA join(875. .944,2644. .2803,3603. .3730,6347. .6548)
/gene="PLA2G1B"
/product="phospholipase A2, group IB (pancreas)"
CDS join(911. .944,2644. .2803,3603. .3730,6347. .6471)
/gene="PLA2G1B"
/codon_start=1
/product="phospholipase A2, group IB (pancreas)"
/protein_id="AAR05441.1"
/db_xref="GI:37953285"
translations="MKLLVAVLLTVAADSGISPRVWQFRKMVKVIGSDPFFLEY
NNYGCYGLSGSGTVDLEDKCCQTHDNCYDQAKLDSCKFLDNPYTHYISCSGS
AITCSSNNKECAPICMCDRNAATCFSPKAPYNKAHKMLDTKKYQCS"
repeat_region /rpt_family="Alu"
1174. .1307
/rpt_type=dispersed
variation 1286
/gene="PLA2G1B"
frequency="0.01"
replace="g"
repeat_region 1319. .1588
/rpt_family="Alu"
/rpt_type=dispersed
variation 1343
/gene="PLA2G1B"
frequency="0.03"
replace="a"
variation 1593
/gene="PLA2G1B"
frequency="0.02"
replace="t"
repeat_region 1605. .1669
/rpt_family="L2"
/rpt_type=dispersed
misc_feature 1638. .2188
/gene="PLA2G1B"
note="Region not scanned for variation"
1670. .1975
/rpt_family="Alu"
/rpt_type=dispersed
repeat_region 1976. .2038
/rpt_family="L2"
/rpt_type=dispersed
repeat_region 2086. .2165
/rpt_family="L2"
/rpt_type=dispersed
repeat_region 2194. .2319
/rpt_family="Alu"
/rpt_type=dispersed
misc_feature 2319. .2453
/gene="PLA2G1B"
note="Region not scanned for variation"
3027
/gene="PLA2G1B"
frequency="0.25"
replace="t"
repeat_region 3165. .3431
/rpt_family="Alu"
/rpt_type=dispersed
variation 3262
/gene="PLA2G1B"
frequency="0.01"
replace="t"
variation 3321
/gene="PLA2G1B"
frequency="0.01"
replace="a"
variation 3396
/gene="PLA2G1B"
frequency="0.03"
replace="g"
variation 3630
/gene="PLA2G1B"
frequency="0.10"
replace="c"
variation 3702
/gene="PLA2G1B"
frequency="0.15"
replace="a"
variation 3768
/gene="PLA2G1B"
frequency="0.01"
replace="t"
variation 3818
/gene="PLA2G1B"
frequency="0.01"
replace="a"
repeat_region 3846. .3964
/rpt_family="MIR"
/rpt_type=dispersed
repeat_region 3977. .4280
/rpt_family="Alu"
/rpt_type=dispersed
repeat_region 4281. .4346
/rpt_family="L2"
/rpt_type=dispersed
repeat_region 4347. .4634
/rpt_family="Alu"
/rpt_type=dispersed
variation 4386
/gene="PLA2G1B"
frequency="0.08"
replace="t"
repeat_region 4635. .4723
/rpt_family="L2"
/rpt_type=dispersed
variation 5013
/gene="PLA2G1B"
frequency="0.10"
replace="a"
repeat_region 5103. .5404
/rpt_family="Alu"
/rpt_type=dispersed
variation 5133
/gene="PLA2G1B"
frequency="0.01"
replace="c"
variation 5543
/gene="PLA2G1B"
frequency="0.01"
replace="c"
variation 5555
/gene="PLA2G1B"
frequency="0.01"
replace="t"
variation 5556
/gene="PLA2G1B"
frequency="0.15"
replace="g"
repeat_region 5577. .5629
/rpt_family="L1"
/rpt_type=dispersed
variation 5580
/gene="PLA2G1B"
frequency="0.09"
replace="g"
variation 5675
/gene="PLA2G1B"
frequency="0.09"
replace="g"
variation 5779. .6083
/rpt_family="Alu"
/rpt_type=dispersed
variation 5929
```

/gene="PLA2G1B"
/frequency="0.01"
/replace="a"
5949
variation
/gene="PLA2G1B"
/frequency="0.01"
/replace="g"
6039
variation
/gene="PLA2G1B"
/frequency="0.01"
/replace="a"
6082
variation
/gene="PLA2G1B"
/frequency="0.14"
/replace="t"

Query Match 100.0%; Score 2499.4; DB 8; Length 8368;
Best Local Similarity 99.8%; Pred. No. 0;
Matches 2497; Conservative 4; Mismatches 0; Indels 0; Gaps 0;
Qy 1 GTCGTGTCACCTGCTGCTGAGCTGGGTAAACAGAGCAACTCTGTCTCAAAAAAATAATG 60
Db 3374 GTCGTGTCACCTGCTGCTGAGCTGGGTAAACAGAGCAACTCTGTCTCAAAAAAATAATG 3433
Qy 61 CTTTCAATAAATAATGATAAAGGACTATATTTTTTCAAGCCATAGGATCATTTCTCC 120
Db 3434 CTTTCAATAAATAATGATAAAGGACTATATTTTTTCAAGCCATAGGATCATTTCTCC 3493
Qy 121 TGAAGCATCTGGCGAAGTCATCCCACTGCTGCTGAGAGTGGCGAGGTGAGGGCTGAC 180
Db 3494 TGAAGCATCTGGCGAAGTCATCCCACTGCTGCTGAGAGTGGCGAGGTGAGGGCTGAC 3553
Qy 181 CTATTGCTGTCACCTACTCTATCTCAGCTGCTCCCTCCCACTTTCCAGGTGCTGCCAGA 240
Db 3554 CTATTGCTGTCACCTACTCTATCTCAGCTGCTCCCTCCCACTTTCCAGGTGCTGCCAGA 3613
Qy 241 CACATGCAACTGCTATGACAGAGCCCAAGAGCTGGAAGCTGTAATTTCTGCTGACA 300
Db 3614 CACATGCAACTGCTATGACAGAGCCCAAGAGCTGGAAGCTGTAATTTCTGCTGACA 3673
Qy 301 MMCGTACACCCACCACTATTACATCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 360
Db 3674 ACCGTTACACCCACCACTATTACATCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 3733
Qy 361 GGTATTATCCCTTCTTGAACCTATGAATTTAGTTGGTTCTCAGTAGGCCGGGGGAATA 420
Db 3734 GGTATTATCCCTTCTTGAACCTATGAATTTAGTTGGTTCTCAGTAGGCCGGGGGAATA 3793
Qy 421 ATAGTAAACACAGCATGATTTAGTGTAAATTTCTGGTTCTGGGCAGTGTCTCCTTTA 480
Db 3794 ATAGTAAACACAGCATGATTTAGTGTAAATTTCTGGTTCTGGGCAGTGTCTCCTTTA 3853
Qy 481 ATCCTGAGAACACACTATGGGATAGTACAAATTTATCTCACTTAAACAGATAGAAACT 540
Db 3854 ATCCTGAGAACACACTATGGGATAGTACAAATTTATCTCACTTAAACAGATAGAAACT 3913
Qy 541 GAGGCTCAGAAGGCTGAGCTATTTGCCCCAAGATCACACAGCTTCTAAGTGTGACAGTTT 600
Db 3914 GAGGCTCAGAAGGCTGAGCTATTTGCCCCAAGATCACACAGCTTCTAAGTGTGACAGTTT 3973
Qy 601 GGGTTTTTTTTTGTGTTTGTAGACAGAGGCTTCTGCTCTGTCACCCAGGCATGAGCAC 660
Db 3974 GGGTTTTTTTTTGTGTTTGTAGACAGAGGCTTCTGCTCTGTCACCCAGGCATGAGCAC 4033
Qy 661 AGTGGTCAACCATAGTCACTGAGCTCAACCTCTCTGAGCTCAAGGGATCTGCTGACC 720
Db 4034 AGTGGTCAACCATAGTCACTGAGCTCAACCTCTCTGAGCTCAAGGGATCTGCTGACC 4093
Qy 721 TCAGCTCCCAAGTGTGGAATTTACGAGCGTGCACCAACCGCTCGCTTAATTAATAA 780
Db 4094 TCAGCTCCCAAGTGTGGAATTTACGAGCGTGCACCAACCGCTCGCTTAATTAATAA 4153
Qy 781 ATTTTTTTGTAGAGACTGGGTCTTACTACGTTGGCCAGGCTTGTCTTAAACTCTCGCTT 840

Db 4154 ATTTTTTTGTAGAGACTGGGTCTTACTACGTTGGCCAGGCTTGTCTTAAACTCTCGCTT 4213
Qy 841 CAAGCAATCTCTCTACCTTGGCATCCAAAGTCTCTGGATTACAGGGGTGAGCCACCATG 900
Db 4214 CAAGCAATCTCTCTACCTTGGCATCCAAAGTCTCTGGATTACAGGGGTGAGCCACCATG 4273
Qy 901 TGGGGCTACTTATTTCTTTTACATTTCCATCTTTTCCAATAGAAATGTAAGATCCACAGAACAG 960
Db 4274 TGGGGCTACTTATTTCTTTTACATTTCCATCTTTTCCAATAGAAATGTAAGATCCACAGAACAG 4333
Qy 961 GGAATTAAGTCTATTTTCTTCTCTTTTGGAGACAGAGTCTCTCATCTCATCACTCAA 1020
Db 4334 GGAATTAAGTCTATTTTCTTCTCTTTTGGAGACAGAGTCTCTCATCTCATCACTCAA 4393
Qy 1021 CCTCGGTTAGCTCAGTCACTGCAACCTCTGCTCCCGGGTTCAAGGATTTCTCTGCTTAAGC 1080
Db 4394 CCTCGGTTAGCTCAGTCACTGCAACCTCTGCTCCCGGGTTCAAGGATTTCTCTGCTTAAGC 4453
Qy 1081 CTCCTGAGTAGCTGGAATTTACAAAGCTGCAACCACTGCTTGGCTAATTTTGTATTTT 1140
Db 4454 CTCCTGAGTAGCTGGAATTTACAAAGCTGCAACCACTGCTTGGCTAATTTTGTATTTT 4513
Qy 1141 TAGCAGAGATGGGGTTTTACCATGTTGCCAGGCTGCTCTCAAACTCCTCACTCAAGTG 1200
Db 4514 TAGCAGAGATGGGGTTTTACCATGTTGCCAGGCTGCTCTCAAACTCCTCACTCAAGTG 4573
Qy 1201 ATCTGCTGCTCAGTCTCCCAAGTCTGGAATTTATAGGCGTGAAGTCACTGCTGCTGCG 1260
Db 4574 ATCTGCTGCTCAGTCTCCCAAGTCTGGAATTTATAGGCGTGAAGTCACTGCTGCTGCG 4633
Qy 1261 CGAATTAAGTCTTATTTCTTTTATGCTATATCCCAAGTCTAGAGAGTGTCTGACATAT 1320
Db 4634 CGAATTAAGTCTTATTTCTTTTATGCTATATCCCAAGTCTAGAGAGTGTCTGACATAT 4693
Qy 1321 AGTAGTGTCTCAATAAATAATTTGATGAATGCAAGCTTAGATATAAATCTTTCTTTTCTT 1380
Db 4694 AGTAGTGTCTCAATAAATAATTTGATGAATGCAAGCTTAGATATAAATCTTTCTTTTCTT 4753
Qy 1381 TTTTTTAAACAATCTTTGCAACTTTGCAAAATAAATACAACTTTGCAATCTGCTTTTCA 1440
Db 4754 TTTTTTAAACAATCTTTGCAACTTTGCAAAATAAATACAACTTTGCAATCTGCTTTTCA 4813
Qy 1441 CTTATCACCTTGTATGACCTTTTTCATATTTGCTTCAAACTTTTATTTGTTACTGTTTTC 1500
Db 4814 CTTATCACCTTGTATGACCTTTTTCATATTTGCTTCAAACTTTTATTTGTTACTGTTTTC 4873
Qy 1501 ATTTGTTACTATTTTGTAGTCACTGAATTAATTTGCTTAAATTTGCTTATACTCTCTGCTC 1560
Db 4874 ATTTGTTACTATTTTGTAGTCACTGAATTAATTTGCTTAAATTTGCTTATACTCTCTGCTC 4933
Qy 1561 CACTTTTAGAGGCCAAATTTTACAAATCTGATGAAGCTATGAACCTCTCTCCAGAGAAA 1620
Db 4934 CACTTTTAGAGGCCAAATTTTACAAATCTGATGAAGCTATGAACCTCTCTCCAGAGAAA 4993
Qy 1621 TACACACACACACACACTCACACAGTTTTTTTTTTAAATTTGTTGCAACTAAGACAAGA 1680
Db 4994 TACACACACACACACACTCACACAGTTTTTTTTTTAAATTTGTTGCAACTAAGACAAGA 5053
Qy 1681 AACCTGCATTTAGAGGATGTTTGTTCATATTTAAATTTAAATAAATCACTAGTTGGGCACAGTGA 1740
Db 5054 AACCTGCATTTAGAGGATGTTTGTTCATATTTAAATTTAAATAAATCACTAGTTGGGCACAGTGA 5113
Qy 1741 CTCAGCCTGTAAACACAGTACTTTGGAAAGTCCAAGGTGGGTGGATCACTTGAAGTGAGA 1800
Db 5114 CTCAGCCTGTAAACACAGTACTTTGGAAAGTCCAAGGTGGGTGGATCACTTGAAGTGAGA 5173
Qy 1801 AGTTTCAGAGCCAGCTGGTCAATATGCTGAAACCTCTATCTCTACTAAAAATACAAAAAT 1860
Db 5174 AGTTTCAGAGCCAGCTGGTCAATATGCTGAAACCTCTATCTCTACTAAAAATACAAAAAT 5233
Qy 1861 AGCTGGGTGTAGTGTGATGCTGCTGTAGTCCAGCTACTCTGGAGGCTGAGGCAAGAGAA 1920

Db 5234 AGCTGGGTGTAGTGCATGCTGTAGTCCAGCTACTCGGAGGCTGAGGCAAGAGAA 5293

Qy 1921 TTGCTTGAACCTGGAGCAGAGGTTGCAGTGAGCCGAGATCCCACCACTGCATCCAGC 1980

Db 5294 TTGCTTGAACCTGGAGCAGAGGTTGCAGTGAGCCGAGATCCCACCACTGCATCCAGC 5353

Qy 1981 CTGGGCGACACAGCGAGACTCTATCTCAAAAAATAAATAAATAAATAAAGGATCGGAG 2040

Db 5354 CTGGGCGACACAGCGAGACTCTATCTCAAAAAATAAATAAATAAATAAAGGATCGGAG 5413

Qy 2041 AGAAACAAAATAAAGATTCCTGAAGGTAAGCAGAGATACGTAATATATATGTAATAA 2100

Db 5414 AGAAACAAAATAAAGATTCCTGAAGGTAAGCAGAGATACGTAATATATATGTAATAA 5473

Qy 2101 AGTTTAATGCATTTTAACCTGTAATCTATTTGTTTATTTTGGTTTATAAAGTAACAAGC 2160

Db 5474 AGTTTAATGCATTTTAACCTGTAATCTATTTGTTTATTTTGGTTTATAAAGTAACAAGC 5533

Qy 2161 CAAAAGTAATCAACTTCAAACTCTACATAAATATCTATTATGAAAAGTGAAGGCATCT 2220

Db 5534 CAAAAGTAATCAACTTCAAACTCTACATAAATATCTATTATGAAAAGTGAAGGCATCT 5593

Qy 2221 ATAATCCTACTACCAAGATAAACAGTTACATATTCCTCAGATTTTGGGGCATACAC 2280

Db 5594 ATAATCCTACTACCAAGATAAACAGTTACATATTCCTCAGATTTTGGGGCATACAC 5653

Qy 2281 TAGCTTTTATTTGGGAAAATTCATGTCGAGGCATACCTAATTTTCTAAATGCT 2340

Db 5654 TAGCTTTTATTTGGGAAAATTCATGTCGAGGCATACCTAATTTTCTAAATGCT 5713

Qy 2341 ATGTAGTATTCATTTAAGGATGTTCCATAATTTTAAATACATGCTTTAAAGTAGAGA 2400

Db 5714 ATGTAGTATTCATTTAAGGATGTTCCATAATTTTAAATACATGCTTTAAAGTAGAGA 5773

Qy 2401 AACTAGTGTGGGCATGGTGGCTCAGCTGATCCAGCACTTTGGGAGGCCGAGGCAAA 2460

Db 5774 AACTAGTGTGGGCATGGTGGCTCAGCTGATCCAGCACTTTGGGAGGCCGAGGCAAA 5833

Qy 2461 TGGATCACTTCAGGTCGGAGTTTGAGCCAGCCTGACAA 2501

Db 5834 TGGATCACTTCAGGTCGGAGTTTGAGCCAGCCTGACAA 5874

RESULT 2

AC003982/2

LOCUS AC003982 122302 bp DNA linear PRI 18-MAR-1999

DEFINITION Homo sapiens PAC clone 166H1 from 12q, complete sequence.

ACCESSION AC003982

VERSION AC003982.1 GI:2769695

KEYWORDS HTG.

SOURCE Homo sapiens (human)

ORGANISM Homo sapiens

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini; Hominidae; Homo.

1 (bases 1 to 122302)

Sulston, J.E. and Waterston, R.

Toward a complete human genome sequence

JOURNAL Genome Res. 8 (11), 1097-1108 (1998)

PUBMED 9847074

REFERENCE 2 (bases 1 to 122302)

Bradshaw, H., Wu, X. and Ozersky, P.

The sequence of Homo sapiens PAC clone 166H1

JOURNAL Unpublished (1999)

REFERENCE 3 (bases 1 to 122302)

Waterston, R.

Direct Submission

TITLE Submitted (13-JAN-1998) Department of Genetics, Washington University, 4444 Forest Park Avenue, St. Louis, Missouri 63108, USA

JOURNAL

REFERENCE 4 (bases 1 to 122302)

Waterston, R.

Direct Submission

TITLE Submitted (18-MAR-1999) Department of Genetics, Washington University, 4444 Forest Park Avenue, St. Louis, Missouri 63108, USA

JOURNAL

COMMENT

University, 4444 Forest Park Avenue, St. Louis, Missouri 63108, USA
SUBMITTED BY: WUGSC
Genome Sequencing Center
Department of Genetics
Washington University
St. Louis MO 63108, USA
http://genome.wustl.edu/gsc
mailto:sapiens@watson.wustl.edu

NOTICE: This sequence may not represent the entire insert of this clone. It may be shorter because we only sequence overlapping clone sections once, or longer because we provide a small overlap between neighboring data submissions.

This sequence was finished as follows unless otherwise noted: all regions were double stranded, sequenced with an alternate chemistry, or covered by high quality data (i.e., phred quality >= 30); an attempt was made to resolve all sequencing problems, such as compressions and repeats; all regions were covered by sequence from more than one subclone; and the assembly was confirmed by restriction digest.

MAPPING INFORMATION:

This clone was originally isolated in the laboratory of Professor Graeme Bell, Howard Hughes Medical Institute and Departments of Biochemistry and Molecular Biology, and Medicine, The University of Chicago, Chicago, IL, USA. The clone was provided by the laboratory of Dr. Roger Cox at The Wellcome Trust Centre For Human Genetics, Oxford, UK. Some contig information was also obtained from Yamagata et al., Nature 384:455-8 (1996).

SOURCE INFORMATION:

This clone was derived from human PAC library RPCI-1, prepared by Pieter de Jong and coworkers at the Roswell Park Cancer Institute (http://baepac.med.buffalo.edu) using the method described by Ioannou et al., Nature Genetics 6:84-9 (1994). The library is from one male donor.
The clone may be obtained either from Genome Systems, Inc. (http://www.genomesystems.com) or Research Genetics, Inc. (http://www.resgen.com); or from Pieter de Jong.

VECTOR: pCYPAC2

NEIGHBORING SEQUENCE INFORMATION:

The clone sequenced to the left is 278C19; the clone sequenced to the right is 15E1. Actual start of this clone is at base position 1 of 166H1; actual end is at 122302 of 166H1.

FEATURES

Source
1. .122302
/organism="Homo sapiens"
/mol_type="genomic DNA"
/db_xref="taxon:9606"
/chromosome="12"
/map="12q"
/clone="166H1"
/clone_lib="RPCI-1"
1. .238
/rpt_family="Alu"
239. .265
/rpt_family="AT_rich"
268. .565
/rpt_family="Alu"
1001. .1309
/rpt_family="Alu"
1440. .1745
/rpt_family="Alu"
1909. .1940
/rpt_family="MER1_type"
1955. .1983
/rpt_family="AT_rich"
2063. .2362
/rpt_family="Alu"
2516. .2650
/rpt_family="Alu"
2674. .2730

repeat_region	11568. .11705	/rpt_family="L1"	11568. .11705	/rpt_family="Alu"
repeat_region	11737. .12032	/rpt_family="GAAAA)n"	11737. .12032	/rpt_family="Alu"
repeat_region	2790. .3077	/rpt_family="Alu"	12034. .12066	/rpt_family="7SLRNA"
repeat_region	3081. .3139	/rpt_family="L1"	12075. .12372	/rpt_family="Alu"
repeat_region	3716. .4020	/rpt_family="Alu"	12448. .12753	/rpt_family="Alu"
repeat_region	4196. .4492	/rpt_family="Alu"	12763. .12874	/rpt_family="L2"
gene	4560. .14034	/gene="WUGSC:H.166H1.1"	12999. .13303	/rpt_family="Alu"
CDS	join(4560. .5056,13418. .13748,13882. .14034)	/genes="WUGSC:H.166H1.1"	complement(13674. .13752)	/notes="match to EST N80020 (NID:g1242721) za91a08.sl"
	/notes="unknown function; 60% similar to Z50177 (PID:g927403) (PID:g927402); H_166H1.1"	/codon_start=1	complement(13840. .14195)	/notes="match to EST N80020 (NID:g1242721) za91a08.sl"
	/evidence=not_experimental	/protein_id="AAB95634.1"	14196. .14229	/rpt_family="AT_rich"
	/db_xref="GI:2769696"	/translation="MKMSFALTPRSAGRWIANPSPCKASIGLFPVAPPLDPKVK	14240. .14339	/rpt_family="U6"
	KELORTILSKLLVGLGISTESGIPDYSEKVLGYARTDRRPIQHGDFVRSAPIR	QRYAANFVGMQFQSHQNPAPHWALSTWEKGLYLVATQNDALHTKAGSRRLTEL	14349. .14646	/rpt_family="Alu"
	HGCMORAYCSVFLSGRLVCLDCGQTERGVLOERFQVLNPTWSAEHGLAPDGVF	LSEQYCFQVPTQVCGGHLKPDVVFQDTPNPKVDFVHKRVKESDILLVVGSSIQ	14652. .14940	/rpt_family="Alu"
	VSYGRFILTAEKKLPALINLIGPTRSDDLACLKLSRCGELLPLIDPC"	5366. .5655	/rpt_family="Alu"	
repeat_region	5672. .5798	/rpt_family="Alu"	100.0%; Score 2499.4; DB 8; Length 122302;	
repeat_region	5802. .5834	/rpt_family="Alu"	Best Local Similarity 99.8%; Pred. No. 0;	
repeat_region	5834. .5857	/rpt_family="(TA)n"	Matches 2497; Conservative 4; Mismatches 0; Indels 0; Gaps 0;	
repeat_region	5870. .6184	/rpt_family="POLY_A"		
repeat_region	6224. .6268	/rpt_family="Alu"		
repeat_region	6269. .6558	/rpt_family="(CA)n"		
repeat_region	6729. .7029	/rpt_family="Alu"		
repeat_region	7051. .7172	/rpt_family="Alu"		
repeat_region	7192. .7500	/rpt_family="Alu"		
repeat_region	7518. .7820	/rpt_family="Alu"		
repeat_region	8207. .8500	/rpt_family="Alu"		
repeat_region	8576. .8878	/rpt_family="Alu"		
repeat_region	9023. .9154	/rpt_family="Alu"		
repeat_region	9165. .9470	/rpt_family="Alu"		
repeat_region	9887. .10188	/rpt_family="Alu"		
repeat_region	10221. .10342	/rpt_family="Alu"		
repeat_region	10408. .10707	/rpt_family="L1"		
repeat_region	10713. .11014	/rpt_family="Alu"		
repeat_region	11020. .11110	/rpt_family="Alu"		
repeat_region	11146. .11386	/rpt_family="Alu"		
repeat_region	11208. .11386	/rpt_family="Alu"		
repeat_region	11391. .11515	/rpt_family="Alu"		

Carron, T.F., Carter, M., Cavazos, S.R., Chacko, J., Chavez, D., Chen, G., Chen, R., Chen, Z., Chowdhry, I., Christopoulos, C., Cleveland, C.D., Cox, C., Coyle, M.D., Dathorne, S.R., David, R., Davila, M.L., Davis, C., Davy-Carroll, L., Dederich, D.A., Delaney, K.R., Delgado, O., Dunn, A.L., Ding, Y., Dinh, H.H., Douthwaite, K.J., Draper, H., Dugan-Rocha, S., Durbin, K.J., Earnhart, C., Edgar, D., Edwards, C.C., Elhaj, C., Escotto, M., Falls, T., Ferraguto, D., Flagg, N., Ford, J., Foster, P., Frantz, P., Gabisi, A., Gao, J., Garcia, A., Garner, T., Garza, N., Gill, R., Gorrell, J.H., Guevara, W., Gunaratne, P., Hale, S., Hamilton, K., Harris, C., Harris, K., Hart, M., Havlak, P., Hawes, A., Hernandez, J., Hernandez, O., Hodgson, A., Hogue, M., Holloway, C., Hollins, B., Honsi, F., Howard, S., Huber, J., Hulyk, S., Hume, J., Jackson, L.E., Jacobson, B., Jia, Y., Johnson, R., Jolivet, S., Joudah, S., Karlsson, E., Kelly, S., Khan, U., King, L., Korvah, J., Kovar, C., Kratovic, J., Kureshi, A., Landry, N., Leal, B., Lewis, L.C., Lewis, L., Li, J., Li, Z., Licharge, O., Lieu, C., Liu, J., Liu, W., Loulseged, H., Lozano, R.J., Lu, X., Lucier, R., Lucier, R., Luna, R., Ma, J., Maheshwari, M., Mapua, P., Martin, R., Martindale, A., Martinez, E., Massey, E., Mawhiney, E., McLeod, M.P., Meador, M., Mei, G., Metzker, M., Miner, G., Miner, Z., Mitchell, T., Mohabbat, K., Morgan, M., Morris, S., Moser, M., Neal, D., Newton, J., Newton, N., Nguyen, A., Nguyen, N., Nguyen, N., Nickerson, E., Nwokenkwo, S., Oguh, M., Okwundu, G., Oragunye, N., Oviedo, R., Face, A., Payton, B., Peery, J., Perez, L., Peters, L., Pickens, R., Primus, E., Pu, L.L., Quiles, M., Ren, Y., Rives, M., Rojas, A., Rojibokan, I., Rolfe, M., Ruiz, S., Savary, G., Scherer, S., Scott, G., Shen, H., Shoostari, N., Sison, I., Sodergren, E., Sonaite, T., Sparks, A., Stanley, H., Stone, H., Sutton, A., Svatek, A., Tabor, P., Tamerisa, A., Tamerisa, K., Tang, H., Tansley, J., Taylor, C., Taylor, T., Telford, B., Thomas, N., Thomas, S., Uman, K., Vasquez, L., Vera, V., Villalon, D., Vinson, R., Wang, Q., Wang, S., Ward-Moore, S., Warren, R., Washington, C., Watlington, S., Williams, G., Williamson, A., Wlezyk, R., Wooden, S., Worley, K., Wu, C., Wu, Y., Wu, Y.F., Zhou, J., Zorrilla, S., Nelson, D., Weinstock, G., and Gibbs, R.

Direct Submission
Unpublished
2 (bases 1 to 220384)
Worley, K.C.
Direct Submission
Submitted (11-AUG-2000) Human Genome Sequencing Center, Department of Molecular and Human Genetics, Baylor College of Medicine, One Baylor Plaza, Houston, TX 77030, USA
3 (bases 1 to 220384)
Worley, K.C.
Direct Submission
Submitted (26-MAR-2002) Human Genome Sequencing Center, Department of Molecular and Human Genetics, Baylor College of Medicine, One Baylor Plaza, Houston, TX 77030, USA
On Mar 26, 2002 this sequence version replaced gi:18449664.
----- Genome Center
Center: Baylor College of Medicine
Center code: BCM
Web site: <http://www.hgsc.bcm.tmc.edu/>
Contact: hgsc-help@bcm.tmc.edu
----- Project Information
Center project name: HCCM
Center clone name: RP11-836M11
----- Summary Statistics
Sequencing vector: Plasmid;
Sequencing vector: M13;
Chemistry: Dye-primer Bodipy; 47% of reads
Chemistry: Dye-terminator Big Dye; 53% of reads
Assembly program: Phrap; version 0.990329
Consensus quality: 234783 bases at least Q40
Consensus quality: 241821 bases at least Q30
Consensus quality: 245671 bases at least Q20
Estimated insert size: 219187; sum-of-contigs estimation
Quality coverage: 8.9x in Q20 bases; sum-of-contigs estimation

* NOTE: Estimated insert size may differ from sequence.length
* (see http://www.hgsc.bcm.tmc.edu/docs/Genbank_draft_data.html).
* NOTE: This is a 'working draft' sequence. It currently

* consists of 11 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.

* 1 2156: contig of 2156 bp in length
* 2157: gap of unknown length
* 2257 7112: contig of 4856 bp in length
* 7113 7212: gap of unknown length
* 7213 13283: contig of 6071 bp in length
* 13284 13383: gap of unknown length
* 13384 20660: contig of 7277 bp in length
* 20661 34473: contig of 13713 bp in length
* 34474 34573: gap of unknown length
* 34574 44394: contig of 9821 bp in length
* 44395 44494: gap of unknown length
* 44495 57126: contig of 12632 bp in length
* 57127 57227: gap of unknown length
* 57228 80266: contig of 23040 bp in length
* 80267 109398: contig of 29032 bp in length
* 109399 149812: contig of 40314 bp in length
* 149813 220384: contig of 70472 bp in length.
* 149813 220384: contig of 70472 bp in length.

FEATURES

source

1. 220384
/organism="Homo sapiens"
/mol_type="genomic DNA"
/db_xref="taxon:9606"
/chromosome="12"
/clone="RP11-836M11"
2157..2256
/estimated_length=unknown
7113..7212
/estimated_length=unknown
13284..13383
/estimated_length=unknown
20661..20760
/estimated_length=unknown
34474..34573
/estimated_length=unknown
43395..44394
/estimated_length=unknown
57127..57226
/estimated_length=unknown
80267..80366
/estimated_length=unknown
109399..109498
/estimated_length=unknown
149813..149912
/estimated_length=unknown

ORIGIN

Query Match 100.0%; Score 2499.4; DB 14; Length 220384;
Best Local Similarity 99.8%; Pred. No. 0;
Matches 2497; Conservative 4; Mismatches 0; Indels 0; Gaps 0;
QY 1 GTCTGTCTACTGCTGTCAGCTGGTAAACAGAGCAACTCTGTCTCAAAAAAATG 60
DB 71511 GTCTGTCTACTGCTGTCAGCTGGTAAACAGAGCAACTCTGTCTCAAAAAAATG 71452
QY 61 CTTTCAATAAATATATGATAAAGGACATTTATTTTCAAGCCATAGGATCATTTCTCC 120
DB 71451 CTTTCAATAAATATATGATAAAGGACATTTATTTTCAAGCCATAGGATCATTTCTCC 71392
QY 121 TGAAGCATCTTGGCGAAGTCATCCCACTGTTCTTGAGAGTGGGAGGTGAGGCTGAC 180
DB 71391 TGAAGCATCTTGGCGAAGTCATCCCACTGTTCTTGAGAGTGGGAGGTGAGGCTGAC 71332

Qy	181	CTATTGCTGCACTTACTCTATCTCAGCTGTCCCTCCACCTTTCCAGGTGCTGCAGA	240		70251	CGATTACTGCTCTATTTCTTTTATTTGCTATATCCCGAGATCTTAGAGCAGTGTCTGACATAT	70192	
Db	71331	CTATTGCTGCACTTACTCTATCTCAGCTGTCCCTCCACCTTTCCAGGTGCTGCAGA	71272		Qy	1321	AGTAGGTGCTCAATAATAATAATTGATGAATGCAAGCTGAGATATAAATCTTTTCTTTTCTTT	1380
Qy	241	CACATGACAACTGCTAYGACCAGGCCAAGAAGCTGGACAGCTGTAAATTTCTGCTGGACA	300		Db	70191	AGTAGGTGCTCAATAATAATAATTGATGAATGCAAGCTGAGATATAAATCTTTTCTTTTCTTT	70132
Db	71271	CACATGACAACTGCTATGACCAGGCCAAGAAGCTGGACAGCTGTAAATTTCTGCTGGACA	71212		Qy	1381	TTTTTAAACAATCTTGCAACCTTTTGCAAGATAAATAACAATCTTGCAATTTCTGCTTTTCA	1440
Qy	301	MMCGGTACACCCACACCTATTATPACTGCTGCTCTGGCTCGGCAATCACTGTAGCAGTA	360		Db	70131	TTTTTAAACAATCTTGCAACCTTTTGCAAGATAAATAACAATCTTGCAATTTCTGCTTTTCA	70072
Db	71211	ACCCTGACACCCACACCTATTATPACTGCTGCTCTGGCTCGGCAATCACTGTAGCAGTA	71152		Qy	1441	CTTATCACCTTGTATTGACTTTTTCATATTGCTCTCAAACTTTATTGTTACTGTTTTTTC	1500
Qy	361	GGTTTATCCCTTCTGACCTATGAATTTCTAGTTGGTTCTCAGTAGGCCGGGGGAAATA	420		Db	70071	CTTATCACCTTGTATTGACTTTTTCATATTGCTCTCAAACTTTATTGTTACTGTTTTTTC	70012
Db	71151	GGTTTATCCCTTCTGACCTATGAATTTCTAGTTGGTTCTCAGTAGGCCGGGGGAAATA	71092		Qy	1501	ATTGTTACTATTATTAGTCACTGAATAATATGGCTTAATTTGCTTTATACATCTCTCTGCTC	1560
Qy	421	ATAGTAACAAACAGCCATGATTTAGTGTAAATTTTCTTGGTTCTGGGCAGTGTCTCTTTTA	480		Db	70011	ATTGTTACTATTATTAGTCACTGAATAATATGGCTTAATTTGCTTTATACATCTCTCTGCTC	69952
Db	71091	ATAGTAACAAACAGCCATGATTTAGTGTAAATTTTCTTGGTTCTGGGCAGTGTCTCTTTTA	71032		Qy	1561	CACTTTGAAGGCCAAATTTTACAAATCTGATGAAGCTATGAACCTCTCTCCCGAGAGAAA	1620
Qy	481	ATCCTCAGAAACAAACACTATGGGATPAGGTACAAATATCTCTCACTTAACAGATAAGAAACT	540		Db	69951	CACTTTGAAGGCCAAATTTTACAAATCTGATGAAGCTATGAACCTCTCTCCCGAGAGAAA	69892
Db	71031	ATCCTCAGAAACAAACACTATGGGATPAGGTACAAATATCTCTCACTTAACAGATAAGAAACT	70972		Qy	1621	TACACACACACACACACTCACAACAGTTTTTTTTTAAATGTTTGCACCTAAAGACAAGA	1680
Qy	541	GAGGCTCAGAGGCTGAGCTATTTGGCCAAAGATCACACAGCTTGTAAAGTGTGACAGTTT	600		Db	69891	TACACACACACACACACTCACAACAGTTTTTTTTTAAATGTTTGCACCTAAAGACAAGA	69832
Db	70971	GAGGCTCAGAGGCTGAGCTATTTGGCCAAAGATCACACAGCTTGTAAAGTGTGACAGTTT	70912		Qy	1681	AACCTGCAATTAGAGATGTTTTGTTTCATATTAATTAATAAATAAATCACTCAGTTGGGCACAGTGA	1740
Qy	601	GGGTTTTTTTTTGTGTTTGTAGAGACAGGGTCTTGCTCTGTCCACCGAGCATGAGCAC	660		Db	69831	AACCTGCAATTAGAGATGTTTTGTTTCATATTAATTAATAAATAAATCACTCAGTTGGGCACAGTGA	69772
Db	70911	GGGTTTTTTTTTGTGTTTGTAGAGACAGGGTCTTGCTCTGTCCACCGAGCATGAGCAC	70852		Qy	1741	CTCAAGCTGTAAACACACAGTACTTTTGGAAAGTCCAAGGTGGGTGGATCAGTTGAGGTGAGA	1800
Qy	661	AGTGGTGCAACCATPAGGTCTACTGAGCTCTCAACCTCTGAGCTCAAGGATCTGCTGACC	720		Db	69771	CTCAAGCTGTAAACACACAGTACTTTTGGAAAGTCCAAGGTGGGTGGATCAGTTGAGGTGAGA	69712
Db	70851	AGTGGTGCAACCATPAGGTCTACTGAGCTCTCAACCTCTGAGCTCAAGGATCTGCTGACC	70792		Qy	1801	AGTTTCGAGACACAGCTGGTCAATATATGTTGAAACCTCTATCTCTACTAAAAATACAAAAATT	1860
Qy	721	TCAGCTCCCAAGTAGCTGGGACTACGAGCGTGCACACACACGCTGGCTTAATTAATAAAA	780		Db	69711	AGTTTCGAGACACAGCTGGTCAATATATGTTGAAACCTCTATCTCTACTAAAAATACAAAAATT	69652
Db	70791	TCAGCTCCCAAGTAGCTGGGACTACGAGCGTGCACACACGCTGGCTTAATTAATAAAA	70732		Qy	1861	AGCTGGGTGTAGTGTAGTATGCTGTAGTCCAGCTACTCGGGAGGCTGAGGCAGAGAAA	1920
Qy	781	ATTTTTTTGTAGAGACTGGGTCTTACTAGCTTGCCAGGGTGTGCTTTAAATCTCTGGCTT	840		Db	69651	AGCTGGGTGTAGTGTAGTATGCTGTAGTCCAGCTACTCGGGAGGCTGAGGCAGAGAAA	69592
Db	70731	ATTTTTTTGTAGAGACTGGGTCTTACTAGCTTGCCAGGGTGTGCTTTAAATCTCTGGCTT	70672		Qy	1921	TTGCTTGAACCTCGGAGGAGAGGTTGCACTGAGCCGAGATCCCAACCACTGCACTCCAGC	1980
Qy	841	CAAGCAATCTCTACTTGGCATCCCAAGATGCTGGGATTAACAGGGGTGAGCCACCATG	900		Db	69591	TTGCTTGAACCTCGGAGGAGAGGTTGCACTGAGCCGAGATCCCAACCACTGCACTCCAGC	69532
Db	70671	CAAGCAATCTCTACTTGGCATCCCAAGATGCTGGGATTAACAGGGGTGAGCCACCATG	70612		Qy	1981	CTGGGCGACACAGCGAGACTCTATCTCAAAAAATAAATAAATAAAGGATCGGAG	2040
Qy	901	TGCGGCTACTTATTTCTTTTACATTTCCATCTTTCCAATAGAAATGAAGATCCACAGAACAG	960		Db	69531	CTGGGCGACACAGCGAGACTCTATCTCAAAAAATAAATAAATAAAGGATCGGAG	69472
Db	70611	TGCGGCTACTTATTTCTTTTACATTTCCATCTTTCCAATAGAAATGAAGATCCACAGAACAG	70552		Qy	2041	AGAAACAAAACTAATAAGATTCCTGAAGGTAAAGGATGAGATACGTAATTAATGTAATAA	2100
Qy	961	GGATTACTGCTATTTTCTCTTTTGTAGACAGAGTCTCACTTCAATCACTCAAC	1020		Db	69471	AGAAACAAAACTAATAAGATTCCTGAAGGTAAAGGATGAGATACGTAATTAATGTAATAA	69412
Db	70551	GGATTACTGCTATTTTCTCTTTTGTAGACAGAGTCTCACTTCAATCACTCAAC	70492		Qy	2101	AGTTTAAATGCAATTTTAACTGTAATCTTATTTGTTTATTTTGGTTATATAAAGTAAACAAGC	2160
Qy	1021	CCTCGGTTCACTGCAACCTCTGCTCCCGGTTCAAGYATTTCTCTGCTCAAGC	1080		Db	69411	AGTTTAAATGCAATTTTAACTGTAATCTTATTTGTTTATTTTGGTTATATAAAGTAAACAAGC	69352
Db	70491	CCTCGGTTCACTGCAACCTCTGCTCCCGGTTCAAGCGATTTCTCTGCTCAAGC	70432		Qy	2161	CAAAAAGTAATGCAACTTCAAACTCTACATAAATATCTATTATGGAAGTGGAGGCACTCT	2220
Qy	1081	CTCCTGAGTAGCTGGAATTAACAGCGTGCAACCATGCTTGGCTAAATTTTGTATTTT	1140		Db	69351	CAAAAAGTAATGCAACTTCAAACTCTACATAAATATCTATTATGGAAGTGGAGGCACTCT	69292
Db	70431	CTCCTGAGTAGCTGGAATTAACAGCGTGCAACCATGCTTGGCTAAATTTTGTATTTT	70372		Qy	2221	ATAATCTCTACTACCCAAAGATAACAGTTACATATTTCTCCAGATTTTGGGGCATACAC	2280
Qy	1141	TAGCAGAGATGGGGTTTTTACCATGTGCCCCAGGCTGGTCTCAAACTCTGCACTCAAGTG	1200		Db	69291	ATAATCTCTACTACCCAAAGATAACAGTTACATATTTCTCCAGATTTTGGGGCATACAC	69232
Db	70371	TAGCAGAGATGGGGTTTTTACCATGTGCCCCAGGCTGGTCTCAAACTCTGCACTCAAGTG	70312		Qy	2281	TAGCTTTTTTATTTGGGAAATTTCCATGTGCGAGGCATACCTAAATTTTCTAAATGTCT	2340
Qy	1201	ATCTGCTGCTCAGTCTCCCAAGATGCTGGAATTAAGGGGTGAGTCACTGTGCTGGC	1260		Db	69231	TAGCTTTTTTATTTGGGAAATTTCCATGTGCGAGGCATACCTAAATTTTCTAAATGTCT	69172
Db	70311	ATCTGCTGCTCAGTCTCCCAAGATGCTGGAATTAAGGGGTGAGTCACTGTGCTGGC	70252		Qy	2341	ATGTAGTATTTCCATTTTAAAGGATGTTCCATAATTTTAAAAATACATGCTTTTAAAGTAGAGA	2400
Qy	1261	CGATTACTGCTATTTTCTTTTATTTGCTATATFCCCGAGATCTTAGAGCAGTGTCTGACATAT	1320					

Db 69171 ATGTAGTATTCATTAAAGGATGTTCCATAATTTTAAATATCATGCTTTAAAGTAGAGA 69112

Qy 2401 AACTAGGTTGGCATTGGTGGCTCAGCGCTGTATCCAGCACTTTGGGAGGCCGAGGCAAA 2460

Db 69111 AACTAGGTTGGCATTGGTGGCTCAGCGCTGTATCCAGCACTTTGGGAGGCCGAGGCAAA 69052

Qy 2461 TGGATCACTTGAAGTCCGGAGTTTGAGACCAAGCCCTGGACAA 2501

Db 69051 TGGATCACTTGAAGTCCGGAGTTTGAGACCAAGCCCTGGACAA 69011

RESULT 4

AX377239 13612 bp DNA linear PAT 18-MAR-2002

LOCUS AX377239 Sequence 1 from Patent WO0212562.

DEFINITION AX377239

ACCESSION AX377239

VERSION AX377239.1 GI:19573528

KEYWORDS

SOURCE Homo sapiens (human)

ORGANISM Homo sapiens

REFERENCE 1

AUTHORS Kazemi, A., Kliem, S.E. and Koshiy, B.

TITLE Haplotypes of the pla2glb gene

JOURNAL Patent: WO 0212562-A 1 14-FEB-2002;

Genaisance Pharmaceuticals, Inc. (US)

FEATURES

source Location/Qualifiers

1. 13612.

/organism="Homo sapiens"

/mol_type="unassigned DNA"

/db_xref="taxon:9606"

variation 3845 /note="PS1: polymorphic base G or A"

variation 3968 /note="PS2: polymorphic base A or G"

variation 6060 /note="PS3: polymorphic base A or G"

variation 6844 /note="PS4: polymorphic base G or A"

variation 9531 /note="PS5: polymorphic base G or A"

ORIGIN

Query Match 100.0%; Score 2499.4; DB 6; Length 13612;

Best Local Similarity 99.8%; Pred. No. 0;

Matches 2496; Conservative 5; Mismatches 0; Indels 0; Gaps 0;

Qy 1 GTGCTGTCACTGTCTCAGCGCTGGGTAACAGAGCAACTCTGTCTCAAAAAAATAATG 60

Db 6516 GTGCTGTCACTGTCTCAGCGCTGGGTAACAGAGCAACTCTGTCTCAAAAAAATAATG 6575

Qy 61 CTTTCAATAAATATATGATAAAGGACTATATTTTTCAAGCCATAGGATCAATTTCTCC 120

Db 6576 CTTTCAATAAATATATGATAAAGGACTATATTTTTCAAGCCATAGGATCAATTTCTCC 6635

Qy 121 TGAAGCATCTTGGGAGTCAATCCCACTGTCTCTGAGAGTGGGCGAGGTGAGCGTGC 180

Db 6636 TGAAGCATCTTGGGAGTCAATCCCACTGTCTCTGAGAGTGGGCGAGGTGAGCGTGC 6695

Qy 181 CTATTGCTGTCACTTACTCTATCTCAGCTGTCCCTCCACATTTCCAGGTGCTGCCAGA 240

Db 6696 CTATTGCTGTCACTTACTCTATCTCAGCTGTCCCTCCACATTTCCAGGTGCTGCCAGA 6755

Qy 241 CACATGACAACTCTAYGACAGCCAGCAAGAGCTGGACAGCTGTAAATTTCTCTGGACA 300

Db 6756 CACATGACAACTCTAYGACAGCCAGCAAGAGCTGGACAGCTGTAAATTTCTCTGGACA 6815

Qy 301 MMCCTGACCCACACCTATTTCATCTCGTGTCTGTGCTGTGGCAATCACCTGTAGCAGTA 360

Db 6816 ACCCGTACCCACACCTATTTCATCTCGTGTCTGTGCTGTGGCAATCACCTGTAGCAGTA 6875

Qy 361 GGTATTACCTCTCTTGACCTATGAATTTCTAGTTGGTTCTCAGTAGGCCCGGGGGAATA 420

Db 6876 GGTATTACCTCTCTTGACCTATGAATTTCTAGTTGGTTCTCAGTAGGCCCGGGGGAATA 6935

Qy 421 ATAGTAACAACAGCCATGATTTAGTGTAAATTTCTTGGTCTGGGCGAGTGTCTCTTTA 480

Db 6936 ATAGTAACAACAGCCATGATTTAGTGTAAATTTCTTGGTCTGGGCGAGTGTCTCTTTA 6995

Qy 481 ATCTCAGAACAACTATGGGATAGTACAATTTATCTCCTAAACAGATTAAGAACT 540

Db 6996 ATCTCAGAACAACTATGGGATAGTACAATTTATCTCCTAAACAGATTAAGAACT 7055

Qy 541 GAGGCTCAGAAAGCTGAGCTATTTGCCAAAGATCACACAGCTTGTAGTGGTGAACGTTT 600

Db 7056 GAGGCTCAGAAAGCTGAGCTATTTGCCAAAGATCACACAGCTTGTAGTGGTGAACGTTT 7115

Qy 601 GGGTTTTTTTGTGTGTGTAGACAGAGGCTTGTCTGTCTCACCCAGGCATGAGCAC 660

Db 7116 GGGTTTTTTTGTGTGTGTAGACAGAGGCTTGTCTGTCTCACCCAGGCATGAGCAC 7175

Qy 661 AGTGTGTCAACCATAGTCTCACTCAGCTCAACCTCTCTGAGCTCAAGGGATCTGCTGACC 720

Db 7176 AGTGTGTCAACCATAGTCTCACTCAGCTCAACCTCTCTGAGCTCAAGGGATCTGCTGACC 7235

Qy 721 TCAGCCTCCCAAGTAGCTGGGACTACGAGCGTGCACCAACCAACGCTGGCTAAATTAATAAAA 780

Db 7236 TCAGCCTCCCAAGTAGCTGGGACTACGAGCGTGCACCAACCAACGCTGGCTAAATTAATAAAA 7295

Qy 781 ATTTTTTGTGAGAGCTGGGTCTTACGTTGGCCAGGCTTGTCTTAAACTCTCTGGCTT 840

Db 7296 ATTTTTTGTGAGAGCTGGGTCTTACGTTGGCCAGGCTTGTCTTAAACTCTCTGGCTT 7355

Qy 841 CRAGCAATCCTCTACCTTGGCATCCAAAGTCTGGGATTAAGGGGTGAGCCACCATG 900

Db 7356 CRAGCAATCCTCTACCTTGGCATCCAAAGTCTGGGATTAAGGGGTGAGCCACCATG 7415

Qy 901 TGGCGTACTTATTTCTTTTACATTTCCATTTTCCAATAGAAATGTAAGATCCACAGAACAG 960

Db 7416 TGGCGTACTTATTTCTTTTACATTTCCATTTTCCAATAGAAATGTAAGATCCACAGAACAG 7475

Qy 961 GGATTAAGTCTATTTCTTCTCTTTTGTGAGACAGAGTCTCACTTCACTCACTCACTCAA 1020

Db 7476 GGATTAAGTCTATTTCTTCTCTTTTGTGAGACAGAGTCTCACTTCACTCACTCACTCAA 7535

Qy 1021 CCTCCGTTGAGTCACTGCAACCTCTGCTCCCGGTTCAAGYATTCTCTCTGCTAAGC 1080

Db 7536 CCTCCGTTGAGTCACTGCAACCTCTGCTCCCGGTTCAAGYATTCTCTCTGCTAAGC 7595

Qy 1081 CTCTGAGTAGCTGGAAATTAAGCGTGCACCACTGCTTGGCTAATTTTGTATTTT 1140

Db 7596 CTCTGAGTAGCTGGAAATTAAGCGTGCACCACTGCTTGGCTAATTTTGTATTTT 7655

Qy 1141 TAGCAGAGATGGGGTTTTTACCATTGTTGCCAGGCTGGTCTCAAACTCCTGACCTCAAGTG 1200

Db 7656 TAGCAGAGATGGGGTTTTTACCATTGTTGCCAGGCTGGTCTCAAACTCCTGACCTCAAGTG 7715

Qy 1201 ATCTGCTGCTCAGTCTCCCAAGTCTGGAAATTAAGCGTGCAGTCACTGCTGCTGCG 1260

Db 7716 ATCTGCTGCTCAGTCTCCCAAGTCTGGAAATTAAGCGTGCAGTCACTGCTGCTGCG 7775

Qy 1261 CGAATTAAGTCTATTTTCTTTTATTTGCTATATCCCAAGTCTAGAGCAGTGTCTGACATAT 1320

Db 7776 CGAATTAAGTCTATTTTCTTTTATTTGCTATATCCCAAGTCTAGAGCAGTGTCTGACATAT 7835

Qy 1321 AGTAGTGTCTCAATAAATAATTAAGTAATGCAAGCTAGATAATAAACTTTCTTTTCTT 1380

Db 7836 AGTAGTGTCTCAATAAATAATTAAGTAATGCAAGCTAGATAATAAACTTTCTTTTCTT 7895

Qy 1381 TTTTTHAAAACATCTTGACAACTTTGACAGAAATAAATAAATCTTGTGATTTCTGCTTTTCA 1440

Db 7896 TTTTTHAAAACATCTTGACAACTTTGACAGAAATAAATAAATCTTGTGATTTCTGCTTTTCA 7955

Qy 1441 CTTATCACCTTGTATGACTTTTTTTCATATTTGCTCAAAACCTTTTATTTGTTACTGTTTTTC 1500

Db	7956	CTTATACCTTGTATGACTTTTTCATATGGCTCAAAACCTTTATTTGTTACTGTTTTTTC	8015
Qy	1501	ATTGTTACTATTTAGTCACTGAATAATATGGCTTAATTTGCTTTATACATCTCTCTGCTC	1560
Db	8016	ATTGTTACTATTTAGTCACTGAATAATATGGCTTAATTTGCTTTATACATCTCTCTGCTC	8075
Qy	1561	CACTTTGAAGGCCAAATTTTACAATCTGATGAAGCTATGAACCCCTCTCCCCAGAGAAA	1620
Db	8076	CACTTTGAAGGCCAAATTTTACAATCTGATGAAGCTATGAACCCCTCTCCCCAGAGAAA	8135
Qy	1621	TACACACACACACACACTCACACACAGTTTTTTTTTAAATGTTTGAACCTAAGACAA	1680
Db	8136	TACACACACACACACACTCACACACAGTTTTTTTTTAAATGTTTGAACCTAAGACAA	8195
Qy	1681	AACCTGATTTAGAGATGTTTGTTCATATTAATTAATAAATACTCAGTTGGGCACTGTA	1740
Db	8196	AACCTGATTTAGAGATGTTTGTTCATATTAATTAATAAATACTCAGTTGGGCACTGTA	8255
Qy	1741	CTCAAGCCTGTAACACACAGTACTTTGGAAGTCCCAAGGTGGGTGGATCACTTGAGTGA	1800
Db	8256	CTCAAGCCTGTAACACACAGTACTTTGGAAGTCCCAAGGTGGGTGGATCACTTGAGTGA	8315
Qy	1801	AGTTTCGAGACAGCCTGCTCAATATGTTGAACCCCTATCTCTACTAAAAATACAAAAAT	1860
Db	8316	AGTTTCGAGACAGCCTGCTCAATATGTTGAACCCCTATCTCTACTAAAAATACAAAAAT	8375
Qy	1861	AGCTGGGTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT	1920
Db	8376	AGCTGGGTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT	8435
Qy	1921	TTGCTTGAACCTGGAGGCGAGGCTTCAGTGAGCCGAGATCCACACCTGCATCTCCAGC	1980
Db	8436	TTGCTTGAACCTGGAGGCGAGGCTTCAGTGAGCCGAGATCCACACCTGCATCTCCAGC	8495
Qy	1981	CTGGGCGACACAGCGAGACTCTATCTCAAAAAATAAATAAATAAATAAATAAATAAATA	2040
Db	8496	CTGGGCGACACAGCGAGACTCTATCTCAAAAAATAAATAAATAAATAAATAAATAAATA	8555
Qy	2041	AGAAACAAATAATATAGATTCCTGAAGTTAAGGATGAGAGATACGTAATATATGTAATA	2100
Db	8556	AGAAACAAATAATATAGATTCCTGAAGTTAAGGATGAGAGATACGTAATATATGTAATA	8615
Qy	2101	AGTTTAAATGATTTTAACTGATCTTATTTGTTTATTTTGGTTTATAAAGTAAACAGC	2160
Db	8616	AGTTTAAATGATTTTAACTGATCTTATTTGTTTATTTTGGTTTATAAAGTAAACAGC	8675
Qy	2161	CAAAAGTAATGCAACTTCAAACTCTACATAAATATCTATTATGGAAGTGAAGGCATCT	2220
Db	8676	CAAAAGTAATGCAACTTCAAACTCTACATAAATATCTATTATGGAAGTGAAGGCATCT	8735
Qy	2221	ATAATCTCTACCAAGATAACACAGTTACATATCTCTCAGATTTTTTGGGGCATACAC	2280
Db	8736	ATAATCTCTACCAAGATAACACAGTTACATATCTCTCAGATTTTTTGGGGCATACAC	8795
Qy	2281	TAGCTTTTATTTTGGGAAATTTCCATGTCGAGGCATACCTAATTTTCTAATGCTCT	2340
Db	8796	TAGCTTTTATTTTGGGAAATTTCCATGTCGAGGCATACCTAATTTTCTAATGCTCT	8855
Qy	2341	ATGTAGTATTTCCATTTAAGGATGTTTCCATATTTTTTAAAAATACATGCTTTAAAGTAGAGA	2400
Db	8856	ATGTAGTATTTCCATTTAAGGATGTTTCCATATTTTTTAAAAATACATGCTTTAAAGTAGAGA	8915
Qy	2401	AACTAGTTGGGATGTTGGCTGACGCTGATCCAGACATTTTGGGAGGCCGAGGCAAA	2460
Db	8916	AACTAGTTGGGATGTTGGCTGACGCTGATCCAGACATTTTGGGAGGCCGAGGCAAA	8975
Qy	2461	TGGATCACTTCAGGTCGGAGTTTGAGACACAGCTTGACAA	2501
Db	8976	TGGATCACTTCAGGTCGGAGTTTGAGACACAGCTTGACAA	9016

AC073930	189729 bp	DNA	linear	HTG 01-MAR-2002
LOCUS	Homo sapiens chromosome 12q clone RP11-144B2, WORKING DRAFT			
DEFINITION	SEQUENCE, 8 unordered pieces.			
AC073930	GI:18958593			
VERSION	HTG; HTGS_PHASE1; HTGS_DRAFT; HTGS_FULLTOP.			
KEYWORDS	Homo sapiens (human)			
SOURCE	Homo sapiens			
ORGANISM	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini; Homnidae; Homo.			
REFERENCE	1 (bases 1 to 189729)			
AUTHORS	Muzny,D.M., Adams,C., Adio-Oduola,B., Ali-osman,F.R., Allen,C., Alsbrooks,S.L., Anaratunge,H.C., Are,J.R., Ayele,M., Banks,T., Barbaria,J., Benton,J., Bimaga,K., Blankenburg,K., Bonnin,D., Bouck,J., Bowie,S., Brieva,M., Brown,E., Brown,M., Bryant,N.P., Buha,C., Burck,P., Burkett,C., Burrell,K.L., Byrd,N.C., Carron,T.F., Carter,M., Cavazos,S.R., Chacko,J., Chavez,D., Chen,G., Chen,R., Chen,Z., Chowdhry,I., Christopoulos,C., Cleveland,C.D., Cox,C., Coyle,M.D., Dathorne,S.R., David,R., Davila,M.L., Davis,C., Davy-Carroll,L., Dederich,D.A., Delaney,K.R., Delgado,O., Denn,A.L., Ding,Y., Dinh,H.H., Douthwaite,K.J., Draper,H., Dugan-Rocha,S., Durbin,K.J., Earnhart,C., Edgar,D., Edwards,C.C., Elhaj,C., Escotto,M., Falls,T., Ferraguto,D., Flagg,N., Ford,J., Foster,P., Frantz,P., Gabisi,A., Gao,J., Garcia,A., Garner,T., Garza,N., Gill,R., Gorrell,J.H., Guevara,W., Gunaratne,P., Hale,S., Hamilton,K., Harris,C., Harris,K., Hart,M., Havlak,P., Hawes,A., Hernandez,J., Hernandez,O., Hodgson,A., Hogues,M., Holloway,C., Hollins,B., Homs,F., Howard,S., Huber,J., Hulyk,S., Hume,J., Jackson,L.E., Jacobson,B., Jia,Y., Johnson,R., Jolivet,S., Joudah,S., Karlsson,E., Kelly,S., Khan,U., King,L., Korvah,J., Kovar,C., Kratovic,J., Kureshi,A., Landry,N., Leal,B., Lewis,L.C., Lewis,L., Li,J., Li,Z., Lichtarge,O., Lieu,C., Liu,J., Liu,W., Louissegh,H., Lozado,R.J., Lu,X., Lucier,A., Lucier,R., Martindale,A., Martinez,E., Maheshwari,M., Mapua,P., Martin,R., Martindale,A., Martinez,E., Massey,E., Mahoney,E., McLeod,M.P., Meador,M., Mei,G., Metzker,M., Miner,G., Miner,Z., Mitchell,T., Mohabbat,K., Morgan,M., Morris,S., Moser,M., Neal,D., Newton,J., Newton,N., Nguyen,A., Nguyen,N., Nguyen,N., Nickerson,E., Nwokwenkwo,S., Oguh,M., Okwuonwu,G., Oragunye,N., Oviedo,R., Pace,A., Payton,B., Peery,J., Perez,L., Peters,L., Pickens,R., Primus,E., Pu,L.L., Quiles,M., Ren,Y., Rives,M., Rojas,A., Rojubokan,I., Rolfe,M., Ruiz,S., Saverly,G., Scherer,S., Scott,G., Shen,H., Shooshtari,N., Siason,I., Sodergren,E., Sonaike,T., Sparks,A., Stanley,H., Stone,H., Sutton,A., Svatek,A., Tabor,P., Tamerisa,A., Tamerisa,K., Tang,H., Tansey,J., Taylor,C., Taylor,T., Telford,B., Thomas,N., Thomas,S., Umaney,K., Vasquez,L., Vera,V., Villalon,D., Vinson,R., Wang,Q., Wang,S., Ward-Moore,S., Warren,R., Washington,C., Watlington,S., Williams,G., Williamson,A., Wleczyk,R., Wooden,S., Worley,K., Wu,C., Wu,Y., Wu,Y.F., Zhou,J., Zorrilla,S., Nelson,D., Weinstock,G. and Gibbs,R.			
TITLE	Direct Submission			
JOURNAL	Unpublished			
REFERENCE	2 (bases 1 to 189729)			
AUTHORS	Worley,K.C.			
TITLE	Direct Submission			
JOURNAL	Submitted (05-JUL-2000) Human Genome Sequencing Center, Department of Molecular and Human Genetics, Baylor College of Medicine, One Baylor Plaza, Houston, TX 77030, USA			
REFERENCE	3 (bases 1 to 189729)			
AUTHORS	Worley,K.C.			
TITLE	Direct Submission			
JOURNAL	Submitted (01-MAR-2002) Human Genome Sequencing Center, Department of Molecular and Human Genetics, Baylor College of Medicine, One Baylor Plaza, Houston, TX 77030, USA			
COMMENT	On Feb 27, 2002 this sequence version replaced gi:14861669. ----- Genome Center Center: Baylor College of Medicine Center code: BCM Web site: http://www.hgsc.bcm.tmc.edu/ Contact: hgsc-help@bcm.tmc.edu			

```

----- Project Information
Center project name: HBWS
Center clone name: RP11-144B2
----- Summary Statistics
Sequencing vector: M13; L08821
Sequencing vector: Plasmid; M77789
Chemistry: Dye-terminator Big Dye; 100% of reads
Assembly program: Phrap; version 0.990329
Consensus quality: 191640 bases at least Q40
Consensus quality: 195135 bases at least Q30
Consensus quality: 19786 bases at least Q20
Estimated insert size: 193340; sum-of-contigs estimation
Quality coverage: 0x in Q20 bases; agarose-fp estimation
Quality coverage: 11.7x in Q20 bases; sum-of-contigs estimation
-----
* NOTE: Estimated insert size may differ from sequence length
      (see http://www.hgsc.bcm.tmc.edu/docs/genbank/draft_data.html).
* NOTE: This is a 'working draft' sequence. It currently
* consists of 8 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.
*
* 1 64773: contig of 64773 bp in length
* 64774 64873: gap of unknown length
* 64874 99097: contig of 34224 bp in length
* 99098 99197: gap of unknown length
* 99198 135331: contig of 36134 bp in length
* 135332 135431: gap of unknown length
* 135432 161344: contig of 25913 bp in length
* 161345 161444: gap of unknown length
* 161445 171231: contig of 9787 bp in length
* 171232 171331: gap of unknown length
* 171332 180922: contig of 9591 bp in length
* 180923 181022: gap of unknown length
* 181023 187271: contig of 6249 bp in length
* 187272 187371: gap of unknown length
* 187372 189729: contig of 2358 bp in length.
* Location/Qualifiers
*   1..189729
*     /organism="Homo sapiens"
*     /mol_type="genomic DNA"
*     /db_xref="taxon:9606"
*     /chromosome="12q"
*     /clone="RP11-144B2"
*     64774..64873
*       /estimated_length=unknown
*     99098..99197
*       /estimated_length=unknown
*     135332..135431
*       /estimated_length=unknown
*     161345..161444
*       /estimated_length=unknown
*     171232..171331
*       /estimated_length=unknown
*     180923..181022
*       /estimated_length=unknown
*     187272..187371
*       /estimated_length=unknown
* ORIGIN
Query Match          84.4%; Score 2110; DB 14; Length 189729;
Best Local Similarity 99.9%; Pred.No. 0;
Matches 2110; Conservative 1; Mismatches 1; Indels 0; Gaps 0;
Qy    390 TAGTTGGTCTCAGTAGCCGCGGGGAAATAATAGTAACAACAGCCATGATTAGTGTTA 449
Db    171332 TAGTTGGTCTCAGTAGCCGCGGGGAAATAATAGTAACAACAGCCATGATTAGTGTTA 1713
Qy    450 ATTTTCCTGGTCTGGSCAGTGTCTCCTTTAATCCTCAGAACCAACTATGGCATAGTTA 509

```

Qy	1590	ATGAAGCTATGACCCCTCTCCCGAGAGAATACACACACACACACACACCTCACACACAG	1649
Db	172532	ATGAAGCTATGAACCCCTCTCCCGAGAGAATACACACACACACACACCTCACACACAG	172591
Qy	1650	TTTTTTTTTAATGTTTGCACCTAAGACAAGAAACCTGCATTAGAGGATGTTTGTTCATAT	1709
Db	172592	TTTTTTTTTAATGTTTGCACCTAAGACAAGAAACCTGCATTAGAGGATGTTTGTTCATAT	172651
Qy	1710	TAATTAATAAATACTAGTTGGGCACAGTGAAGTCAAGCCCTGTAAACACAGTACTTTGGAA	1769
Db	172652	TAATTAATAAATACTAGTTGGGCACAGTGAAGTCAAGCCCTGTAAACACAGTACTTTGGAA	172711
Qy	1770	GTCCAAGGTGGTGGATCACTTGAGGTGAGAAATTCAGACACAGCCCTGGTCAATATGGTG	1829
Db	172712	GTCCAAGGTGGTGGATCACTTGAGGTGAGAAATTCAGACACAGCCCTGGTCAATATGGTG	172771
Qy	1830	AAACCCCTATCTTACTAAAAATACAAAAATTTAGCTGGGTGATGATGCATGCCCTGATG	1889
Db	172772	AAACCCCTATCTTACTAAAAATACAAAAATTTAGCTGGGTGATGATGCATGCCCTGATG	172831
Qy	1890	CCAGCTACTCGGAGGCTGAGGCAAGAAATTCCTTGAACCTGGGAGGCGAGAGTTGCA	1949
Db	172832	CCAGCTACTCGGAGGCTGAGGCAAGAAATTCCTTGAACCTGGGAGGCGAGAGTTGCA	172891
Qy	1950	GTGAGCGGAGATCCACCACTGCACCTCCAGCTGGGCGACACACGAGACTCTATCTCAA	2009
Db	172892	GTGAGCGGAGATCCACCACTGCACCTCCAGCTGGGCGACACACGAGACTCTATCTCAA	172951
Qy	2010	AAAAATAAATAAATAAAGGATCGGAGAGAAACAAACTAATAAGATTCCTGGAAGG	2069
Db	172952	AAAAATAAATAAATAAAGGATCGGAGAGAAACAAACTAATAAGATTCCTGGAAGG	173011
Qy	2070	TAACGAGAGATACCTAAATATATATGTAATAAGTTTAAATGCAATTTAACTGTAATCTTA	2129
Db	173012	TAACGAGAGATACCTAAATATATATGTAATAAGTTTAAATGCAATTTAACTGTAATCTTA	173071
Qy	2130	TTGTTTATTTTGGTTATTAAGTAACCAAGCCAAAGTAATGCACTTCAAACTCTACAT	2189
Db	173072	TTGTTTATTTTGGTTATTAAGTAACCAAGCCAAAGTAATGCACTTCAAACTCTACAT	173131
Qy	2190	AAATATCTATTATGGAAGTGAAGGATCTATAATCCTACTACCCAAAGATACCAAGTT	2249
Db	173132	AAATATCTATTATGGAAGTGAAGGATCTATAATCCTACTACCCAAAGATACCAAGTT	173191
Qy	2250	ACATATCTCTCCAGATTTTGGGCGATACACTAGCTTTTATTTTGGGAAAATTTCCAT	2309
Db	173192	ACATATCTCTCCAGATTTTGGGCGATACACTAGCTTTTATTTTGGGAAAATTTCCAT	173251
Qy	2310	GTGAGGCGATACCTAAATTTTCTAAATGCTATCTAGTATTTCCATTTAAGATGTTCCAT	2369
Db	173252	GTGAGGCGATACCTAAATTTTCTAAATGCTATCTAGTATTTCCATTTAAGATGTTCCAT	173311
Qy	2370	AAATTTTAAATAACATGCTTTAAAGTAGAGAAACTAGGTTGGGCACTGGTGCTCACGCCT	2429
Db	173312	AAATTTTAAATAACATGCTTTAAAGTAGAGAAACTAGGTTGGGCACTGGTGCTCACGCCT	173371
Qy	2430	GTATCCGAGCACTTTGGAGGCCGAGGCAATGATCACTTGAGGTCGGAGTTTGAGAC	2489
Db	173372	GTATCCGAGCACTTTGGAGGCCGAGGCAATGATCACTTGAGGTCGGAGTTTGAGAC	173431
Qy	2490	CAGCCTGGACAA 2501	
Db	173432	CAGCCTGGACAA 173443	

RESULT 6
AC068981/c AC068981 178887 bp DNA linear HTG 01-SEP-2000
DEFINITION Homo sapiens chromosome 15 clone RP11-58017, WORKING DRAFT
SEQUENCE, 37 unordered pieces.
ACCESSION AC068981
VERSION AC068981.3 GI:9958291

KEYWORDS HTG; HTGS PHASE1; HTGS_DRAFT.
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
REFERENCE 1 (bases 1 to 178887)
AUTHORS Waterston,R.H.
TITLE The sequence of Homo sapiens clone
JOURNAL Unpublished
REFERENCE 2 (bases 1 to 178887)
AUTHORS Waterston,R.H.
TITLE Direct Submission
JOURNAL Submitted (15-MAY-2000) Genome Sequencing Center, Washington
University School of Medicine, 4444 Forest Park Parkway, St. Louis,
MO 63108, USA
COMMENT On Sep 1, 2000 this sequence version replaced gi:8748938.

----- Genome Center -----
Center: Washington University Genome Sequencing Center
Center code: WUGSC
Web site: http://genome.wustl.edu/gsc/index.shtml
----- Project Information -----
Center project name: H NH0580107
----- Summary Statistics -----
Sequencing vector: M13; 100%
Sequencing vector: Plasmid; 0%
Chemistry: Dye-primed; 100% of reads
Chemistry: Dye-terminator Big Dye; 0% of reads
Assembly program: Phrap; version 0.990319
Consensus quality: 157045 bases at least Q40
Consensus quality: 163476 bases at least Q30
Consensus quality: 167183 bases at least Q20
Insert size: 152000; agarose-fp
Insert size: 174828; sum-of-contigs
Quality coverage: 3.01 in Q20 bases; agarose-fp
Quality coverage: 3.40 in Q20 bases; sum-of-contigs

* NOTE: This is a 'working draft' sequence. It currently
* consists of 37 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.

1 1242: contig of 1242 bp in length
* 1243: gap of unknown length
* 1343: contig of 1093 bp in length
* 2436: gap of unknown length
* 2535: gap of unknown length
* 2644: contig of 1129 bp in length
* 3764: gap of unknown length
* 4864: contig of 1100 bp in length
* 4964: gap of unknown length
* 7695: contig of 2731 bp in length
* 7796: gap of unknown length
* 10187: contig of 2392 bp in length
* 10287: gap of unknown length
* 10288: contig of 2016 bp in length
* 12304: gap of unknown length
* 12403: contig of 1987 bp in length
* 14391: gap of unknown length
* 14391: contig of 1795 bp in length
* 16286: gap of unknown length
* 16386: contig of 1979 bp in length
* 18365: gap of unknown length
* 18464: contig of 2330 bp in length
* 20794: gap of unknown length
* 20894: contig of 2579 bp in length
* 20895: gap of unknown length
* 23474: contig of 1814 bp in length
* 23573: gap of unknown length
* 25387: gap of unknown length
* 25487: contig of 3383 bp in length
* 25488

Qy	900	GTGGGGTACTATTATTTCTTTACATTCCTTTCCATAGATGATCCAGACA	959
Db	90323	GCCTAACTATTTAATGTGAATTTGTATGTGGCATATAAAATCCCTAGCAGTTTGT	90264
Qy	960	GGGATTACTGCTATTTCTTCCTCTCTTTTGGAGACAGATCTCACTTCATCACCTCA	1019
Db	90263	TTTGTGTT-TTGTGTTTTTGGAGACAGATTTTGCTCTGTGGACATGCCACTGGAGTGCAGTG	90205
Qy	1020	ACCTCCGTTGAGTCACTGCAACCTCTGCGCTCCCGGTTCAAGYATTTCTCTGCTCAAG	1079
Db	90204	GCTCAATCTGGCTCACTGCAACCTCCGCTCCCGGTTCAAGCGATTTCTCTGCTCAG	90145
Qy	1080	CCTCTCAGTACGCTGAATACAGCGTGCACACCATGCTTGCTTAATTTTGTATTT	1139
Db	90144	CCTCTCAGTACGCTGAATACAGCGTGCACACCATGCTTGCTTAATTTTGTATTT	90085
Qy	1140	TTAGCAGATGGGGTTTTTACCATGTTGCCAGGCTGGTCTCAAACTCTGACCTCAAGT	1199
Db	90084	TTGCTAGTCAAGTTTAACTATGTTGGCAGGCTGTTTTCGACTCTGACTCAGGT	90025
Qy	1200	GATCTGCTGCTCAGTCTCCCAAGTCTGGAATTAAGCGTGAATCACTGCTGCTGG	1259
Db	90024	GATCCACTGCTCAGTCTCCCAAGTCTGGAATTAAGCGTGAATCACTGCTGCTGG	89965
Qy	1260	CCGATTACTGCTATTTCTTTATGCTATATCCAGATCTAGACAGTCTGACATA	1319
Db	89964	CCAACTAACTTTTCAATTAAGTGAACCTCTAACACACAGCTTATATTAAGTTT	89913
Qy	1320	TAGTAGTGTCTCAATAATAATGATGAATGCACAGGCTAGATATAAACTTTCTTTTCT	1379
Db	89912	-----TGTTTTTTTTTAGACAGAGGTTCTGCTCTCCACCCAGCTTGAG	89867
Qy	1380	TTTTTTAAACAATCTTGACAACCTTTGAGAATAATAATCAATCTTGCAATTTCTTTTC	1439
Db	89866	TGCGCTGCGCAATCATAGTCTCACTAGGACTCAA-----CTGA	89828
Qy	1440	ACTTATCACCTTGTTATGACTTTTTTCATATGCTCTCAACCTTTATTTGTTACTTTTTT	1499
Db	89827	TCCTCTTGCTCTCTGAGTAACTGGGACTACAGGCTTGAGACACCATCCAGCAGAAATAG	89768
Qy	1500	CATTGTTACTATTTAGTCACTGAATAATATGGCTTAATTTGCTTATACATCTCTCTGCT	1559
Db	89767	CTCTCGATTTGCTTTGGCTTAATCACATTTCCACCTTGATGTACTCAAAATGTTTTT	89708
Qy	1560	CCACTTTAGAGGCCAAATTTACAAATCTGATGAAGCTATGAACCTCTCTCCACAGAA	1619
Db	89707	ATTAGTCAATA-----AATTTTCTAATCAATTTGTGCTCACTACAATTTATTTCTATCA	89653
Qy	1620	ATACACACACACACACACTCACACACAGTTTTTTTTTTTAAATGTTTGCACCTAAGACAAG	1679
Db	89652	CAATTTACAGACTTTCAAAATTTGATTTGGTTTTTCTACTGAATAAATCGTAGAGAAAG	89593
Qy	1680	AAACCTGCA-----TTAGAGGATGTTTGTTCATATTAATTAATAAATACTCACTGGGCACA	1736
Db	89592	AACACAGTAACTAAATGATGGAATTCATCCCTTAAACAAATTTACGCCGGCAGC	89533
Qy	1737	GTGACTCAAGCTGTAAACCACTGATCTTTGAAAGTCCAAAGTGGGTGATCACTTTGAGGT	1796
Db	89532	TTGGCTCAGGCTGTAAATCTCAGCACTTTGGGAGGCCAAAGGTAGTGGATCAC--GAGGT	89475
Qy	1797	GAGAACTCGAGACCGCTGGTCAATATGTTGAACCTCTATCTACTAAATATACAAA	1856
Db	89474	CAGGGCTCGAGACAGAGCTACCAACAATGTTGAACCCCGCTCTACTAAAAATACAAA	89415
Qy	1857	AATTAGCTGGGTAGTATGATGATGCTCTGATGCTCTGATGCTCTGATGCTGATGCTGAGG	1916
Db	89414	AATTAGCTGGGTAGTGGGGGCGCTGTTAATACAGTACTCAGGAGGCTGAGGAGG	89355
Qy	1917	AGAAATTGCTGAACCTGGAGGCGAGAGTTGCAAGTCCAGTCCAGATCCACCACTGCACTC	1976
Db	89354	AGAAATTGCTGAACCCGGGAGCGGAAGCTGCAGTGCAGTGCAGTGCAGTGCAGTGCAGTGC	89295

Qy	1977	CAGCTGGCGGACACAGCAGAGCTCTATCTCAAAAAATAAAATAAAATA 2029	
Db	89294	CAGCTGGCGGACACAGCAGAGCTCGTCTCAAAAAACAATAAAAAACA 89242	
LOCUS	AC009244	190162 bp DNA linear PRI 26-MAY-2002	
DEFINITION	Homo sapiens chromosome 7 clone RP11-305M3, complete sequence.		
ACCESSION	AC009244		
VERSION	AC009244.25	GI:21217399	
KEYWORDS	HTG.		
SOURCE	Homo sapiens (human)		
ORGANISM	Homo sapiens		
REFERENCE	1 (bases 1 to 190162)		
AUTHORS	Kaul,R.K., Olson,M.V., Zhou,Y., James,R.A., Rouse,G., Wu,Z., Saenphimmachak,C., Phelps,K.A., Buckley,D., Kibukawa,M., Raymond,C. and Haugen,E.D.		
TITLE	Direct Submission		
JOURNAL	Unpublished		
REFERENCE	2 (bases 1 to 190162)		
AUTHORS	Bubb,K.L., Desmarais,C.L., Ramsey,S.A. and Hubley,R.M.		
TITLE	Direct Submission		
JOURNAL	Submitted (07-AUG-1999) Human Genome Center, University of Washington, Box 352145, Seattle, WA 98195, USA		
REFERENCE	3 (bases 1 to 190162)		
AUTHORS	Kaul,R.K., Zhou,Y., James,R.A., Raymond,C., Haugen,E.D. and Olson,M.V.		
TITLE	Direct Submission		
JOURNAL	Submitted (30-OCT-2000) Genome Center, University of Washington, Box 352145, Seattle, WA 98195, USA		
REFERENCE	4 (bases 1 to 190162)		
AUTHORS	Kaul,R.K., Olson,M.V., Zhou,Y., James,R.A., Rouse,G., Wu,Z., Saenphimmachak,C., Phelps,K.A., Buckley,D., Kibukawa,M., Raymond,C. and Haugen,E.D.		
TITLE	Direct Submission		
JOURNAL	Submitted (26-MAY-2002) Genome Center, University of Washington, Box 352145, Seattle, WA 98195, USA		
COMMENT	On May 26, 2002 this sequence version replaced gi:11038533.		
	----- Genome Center		
	Center: University of Washington Genome Center		
	Center Code: UWGC		
	Web site: http://www.genome.washington.edu		
	Contact: uwgctgs@u.washington.edu		
	----- Project Information		
	Center project name: chr-7		
	Center clone name: RP11-305M3 (djs195)		
	----- Summary Statistics		
	Assembly program: Phrap; version 0.990319		
	Consensus quality: 189969 bases at least Q40		
	Consensus quality: 190110 bases at least Q30		
	Consensus quality: 190161 bases at least Q20		
	Insert size: 190162; sum-of-contigs		
	Quality coverage: 9.7x in Q20 bases; sum-of-contigs		

	Overlapping Sequences:		
	5': RP11-382M23 (UWGC:djs734g) AC093149 1340-bp overlap		
	3': RP11-448A19 (UWGC:djs705) AC078846 8765-bp overlap		

	Sequence Quality Assessment:		
	This entry has been annotated with sequence quality estimates computed by the Phrap assembly program.		
	All manually edited bases have been reduced to quality zero.		
	Quality levels above 40 are expected to have less than 1 error in 10,000 bp.		
	Base-by-base quality values are not generally visible from the Genbank flat file format but are available as part of this entry's ASN.1 file.		
	This sequence was finished as follows unless otherwise noted:		

all regions were either double-stranded or sequenced with an alternate chemistry or covered by high quality data (i.e., Phred quality > 30); an attempt was made to resolve all sequencing problems, such as compressions and repeats; all regions were covered by at least one plasmid subclone or more than one M13 subclone; and the assembly was confirmed by restriction digest.

Sequence Validation:

This sequence has been validated by Multiple Complete Digest fingerprinting. Comparison of the experimentally derived digest fragments with sequence-predicted fragments is given below. The electronically digested sequence consists of both insert and vector, in order to accurately represent the entire circular BAC. Small fragments below a variable cutoff (approximately 400-800 bp) are not resolved in the fingerprint and hence do not appear in the table. There are no significant remaining discrepancies between the experimental and predicted values. Uniquely ordered fragments are separated by dashed lines.

NsiI				BglII				EcoRI			
SeqDerMap	FngPrnt	SeqDerMap	FngPrnt	SeqDerMap	FngPrnt	SeqDerMap	FngPrnt	SeqDerMap	FngPrnt	SeqDerMap	FngPrnt
21251	20929	12417	12174	8696	8794						
12148	11918	2067	2041	6	<800						
199	<800	5666	5826	1775	1773						
3275	3221	254	<800	2000	1991						
20379	20929	3818	3836	6313	6324						
15555	15150	747	735	3256	3226						
675	<800	9196	9237	325	<800						
5493	5475	4154	4101	12989	12723						
516	<800	5820	5826	5955	5951						
5810	5831	118	<800	611	<800						
9488	9435	2524	2359	1995	1991						
2908	2935	8844	8768	11945	12009						
779	840	332	<800	5577	5593						
821	840	564	<800	4139	4124						
315	<800	26	<800	955	940						
3640	3671	4675	4644	2961	2940						
428	<800	1801	1803	5264	5205						
1144	1119	117	<800	1535	1534						
4684	4717	6797	6799	3313	3383						
20	<800	3146	3177	3760	3761						
492	<800	2376	2359	3061	3110						
10935	10879	210	<800	9605	9505						
5921	6010	5135	5110	335	<800						
529	<800	744	735	2190	2194						
4354	4384	881	947	5117	5205						

Query Match 18.5%; Score 462.6; DB 8; Length 190162;
Best Local Similarity 61.3%; Pred. No. 6.9e-83;
Matches 900; Conservative 1; Mismatches 520; Indels 46; Gaps 8;
QY 604 TTTTCTTTTGTGTTTGTAGAGACAGGCTCTTCTCTCACCAGGATGAGCAGT 663
DB 31687 TTTTCTTTTAAAGTAGAACCGAGTCTTGTGGTACCAGGCTGGGGTCACT 31746
QY 664 GTGCAACCATAGGTCATCTGACGCTCAACCTCTGAGCTCAAGGATCTGTGACCTCA 723
DB 31747 AGTGCAGTCGTGTTTCACTGCATCCTTCAATTCCTAGGAACAGGTGATCCTCCACTTCA 31806

```
QY 724 GCCTCCCAAGTAGCTGGGACTAGAGCGTGCAACACACCGCTGGCTTAATTAATAAATTT 783
Db 31807 GTCTTCTGAGTAGCTGGGACTTCAGGCACTGTGCCACCATGCCAGCTAATTTAAAAA 31866
QY 784 TTTTGTAGAGACTGGGCTTACTACTGTGGCCAGGCTGTCTTAAACTCTCGCTCTCAA 843
Db 31867 TGTGTGAAGACACAGGGCTTGTGTGTGTCTAGGCTGTCTCAAAACAGCGAGCTCAA 31926
QY 844 GCAATCTCTCTACCTTGGCATCCCAAGTGTGGATTACAGGGGTGAGCCACCATGTGC 903
Db 31927 GTGATCTCTCTGCTCATCTCTCCCAAGTGTGGATTACAGGTGTGAGCCACCATGCC 31986
QY 904 GGCTACTTATTTCTTTTACATCTCTTCAATAGATGTAGATCCACAGACAGGA 963
Db 31987 AGCCCATTTCTTTTCTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 32046
QY 964 TTAGTGTCTATTTTCTTTCTTTTCTTTTCTTTTCTTTTCTTTTCTTTTCTTTTCTTT 1022
Db 32047 TATTTTATTTTATTTATTTATTTATTTTATTTTATTTTATTTTATTTTATTTT 32106
QY 1023 TCCGT-----TCAGTCACTGCAACTCTGCTCCCGGGTTCAGAGYATTC 1068
Db 32107 GGAGTGCAGTGGCATGATCTCGGCTCACTGAAACCTCTGCTCTCCAGGTTCAAGCGATTC 32166
QY 1069 TCCTGCTTAAGCCTCCTGAGTGTGGAATTACAGCGTGCACACCATGCTTGGCTAAT 1128
Db 32167 TCCTGCTCAAGCCTCCTGAGTGTGGAATTACAGGTGCCTTCCACTATACCGGCTAA- 32225
QY 1129 TTTTGTATTTTATAGCAGATGGGTTTACCATGTGTCAGGCTGGTCTCAAACTCC 1188
Db 32226 TTTTGTATTTTATAGTAGAGCGGGTTTACCATGTAGTGTAGGCTGGTCTCGAATCTCC 32285
QY 1189 TGACCTCAAGTGTCTGCTGCTCAGTCTCCAAAGTGTGGAATTTATAGGCTGAGTC 1248
Db 32286 TGACCTCAGATGATCTGCTGCTCGGCTCCCAAGTGTGGGATTTACAGGTGTGAGCC 32345
QY 1249 ACTGTGCTGGCCGATTAAGTGTCTATTTT---CTTTATGTCTATATCCCGAGTCTAGAG 1305
Db 32346 ACCGACCGGCTGCCCATTTTTTTTTTAAACAAATGCTTACTGACCTACGCTAAGAG 32405
QY 1306 CAGTGTCTGACATATAGTAGTGTCTCAATAAATTTGATGAATGCAGCCTAGATATA 1365
Db 32406 CTGTAAGGCGAAATGAGACAATGTCTTACCCTCAGATCTTATATTTTGTGTGAGATAAT 32465
QY 1366 AACTTTCTTTTCTTTTAAACAAATCTTTGACAACTTTGAGAAATAAATACAACTTG 1425
Db 32466 ACACATATATAAATAATGATGATACAGGACATATATAAATAATGCCATGACAGATTATA 32525
QY 1426 CATTCGTCTTTTTCAC--TTATCACCCTTGTATGACTTTTTCATATTCGCTCAAACTTTA 1484
Db 32526 AATCCATATTCACATACACAAACAGAGGCTTAGGAAAGTTAGATTTCCCACTTTGGGAAG 32585
QY 1485 TTGTTACTGTTTTTCTTCTTACTATTTTGTACCTGAATAATATGCTTAATTTGCTTT 1544
Db 32586 CTGAGGAGGTGGATCACCTGAGTCAAGGATTTCA--AGACCAGGTGGCCACATGGTG 32643
QY 1545 ATACATCTCTGCTTCACTTTTGAAGGCCAAATTTTCAAACTGTGATGAAGCTATGAAC 1604
Db 32644 AAACCTCTGCTTTACTAAGAGTACAAAAATTAGCTGAGTGTGGTGTGCGAGGCTGTAAT 32703
QY 1605 CCTCTCCCCAGAGAAATACACACACACACACTCACACAGCTTCTTTTAAATGTT 1664
Db 32704 CCTAGCTACTCATGAGGCTGAGGAGGAGAAATCGCTTGAACCCAGGAGGCGGCTGCA 32763
QY 1665 TGCAACTAAGACAGAAACCTTCATAGAGGATGTTTG----- 1702
Db 32764 GTGAGCCGAGATCGGCCACTGCACTCTAGCTTGGGTGACAGATGAGACTCCATCTCAA 32823
QY 1703 TTCAATTAATTAATAAATACTAGTTGGGACAGTGTGCTCAAGCCTGTAAACACAGTAC 1762
Db 32824 AAAAAAAAAAAAAAAAAAGAGCCAGGCGGTGGCTCACACCTGTAAATCCAGCAC 32883
```

```
QY 1763 TTTGGAAGTCCAAGGTGGGTGGATCACTTGAGTGAAGATTTCGAGACAGCCTGGTCAA 1822
Db 32884 TTTGGAGAGCTGAGGTGGGCGGATCA--TGAGTCAAGATTTCGAGACAGCCTGGCCAA 32941
QY 1823 TATGTTGAACCCCTATCTCTACTATAAAATACAAAAATTAGTGGGTGTAGTGTGATGC 1882
Db 32942 TATGTTGAACCCCGTCTTTTACTTAAAGTACAAAAATTAGTCTGGGCGTGTGTCAGGCAC 33001
QY 1883 CTGTAGTCCAGTACTTCGGGAGGCTGAGGCAAGAGAAATGCTTGAACCTGGGAGGCAGA 1942
Db 33002 CTGTAGTCCAGTACTTCAGGAGGCTTAAGGCAGGAGAAATCGCTTGAACCCAGGAGCAGA 33061
QY 1943 GGTTCAGTGTAGCCGAGATCCCACTGCTCCAGCTGGGCGCACACAGCAGACTCT 2002
Db 33062 GGTTCAGTGTAGCCGAGATTTATGCCACTGCACTCCAGCTGGGTGACAGCAAGACTCC 33121
QY 2003 ATCTCAAAAAATAAATAAATAAATA 2029
Db 33122 GTCTCAAAAAAATAAAGAAAGAAAA 33148

RESULT 8
AP000589
LOCUS
DEFINITION
  Homo sapiens chromosome 11 clone CMB9-105N5 map 11q13, WORKING
  DRAFT SEQUENCE, 19 unordered pieces.
ACCESSION
  AP000589.3 GI:8118795
VERSION
  HTG; HTGS_PHASE1; HTGS_DRAFT.
KEYWORDS
  Homo sapiens (human)
SOURCE
  Homo sapiens
  Organism
    Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini;
    Hominidae; Homo.
REFERENCE
  1 (bases 1 to 123192)
    Hattori,M., Ishii,K., Toyoda,A., Taylor,T.D., Hong-Seog,P.,
    Fujiyama,A., Yada,T., Totoki,Y., Watanabe,H. and Sakaki,Y.
    Homo sapiens 123,192 genomic DNA of 11q13
  2 (bases 1 to 123192)
    Hattori,M., Ishii,K., Toyoda,A., Taylor,T.D., Hong-Seog,P.,
    Fujiyama,A., Yada,T., Totoki,Y., Watanabe,H. and Sakaki,Y.
    Direct Submission
    Submitted (12-OCT-1999) Masahira Hattori, The Institute of Physical
    and Chemical Research (RIKEN), Genomic Sciences Center (GSC);
    Kitasato Univ., 1-15-1 Kitasato, Sagamihara, Kanagawa 228-8555,
    Japan (E-mail:hattori@gsc.riken.go.jp,
    URL:http://hgp.gsc.riken.go.jp/, Tel:81-42-778-9923,
    Fax:81-42-778-9924)
    On May 31, 2000 this sequence version replaced gi:6997479.

COMMENT
  ----- Genome Center
  Center: RIKEN Genomic Sciences Center (GSC)
  Center code: RIKEN
  Web site: http://hgp.gsc.riken.go.jp/
  Contact: hattori@gsc.riken.go.jp
  ----- Project Information
  Center project name: HumDraft11
  Center clone name: CMB9-105N5
  ----- Summary Statistics
  Sequencing vector: PCR products; 100% of reads
  Chemistry: Dye-terminator ET-amersham; 100% of reads
  Assembly program: Phrap; version 0.990329
  Consensus quality: 115116 bases at least Q40
  Consensus quality: 118340 bases at least Q30
  Consensus quality: 120191 bases at least Q20
  Insert size: 121392; sum-of-contigs
  Quality coverage: 5.08x in Q20 bases; sum-of-contigs
  -----
  NOTE: This is a 'working draft' sequence. It currently consists of
  19 contigs. The true order of the pieces is not known and their
  order in this sequence record is arbitrary. Gaps between the
  contigs are represented as runs N, but the exact sizes of the gaps
  are unknown. This record will be updated with the finished sequence
```

as soon as it is available and the accession number will be preserved

```

1 15827 contig of 15827 bp in length
  15928 contig of 14557 bp in length
  30585 contig of 10018 bp in length
  40602 contig of 11210 bp in length
  51912 contig of 8010 bp in length
  52013 contig of 10316 bp in length
  60123 contig of 7407 bp in length
  70539 contig of 7672 bp in length
  78046 contig of 6582 bp in length
  85818 contig of 5716 bp in length
  92500 contig of 5833 bp in length
  98316 contig of 4918 bp in length
  104249 contig of 3191 bp in length
  109267 contig of 2519 bp in length
  112558 contig of 1887 bp in length
  115177 contig of 1489 bp in length
  117164 contig of 1673 bp in length
  118753 contig of 1673 bp in length
  120526 contig of 1318 bp in length
  121944 contig of 1249 bp in length
  123192 contig of 1249 bp in length

```

Sequence updated (18-Oct-1999)

Sequence updated (26-May-2000).

- * NOTE: This is a 'working draft' sequence. It currently consists of 19 contigs. The true order of the pieces is not known and their order in this sequence record is arbitrary. Gaps between the contigs are represented as runs of N, but the exact sizes of the gaps are unknown.
- * This record will be updated with the finished sequence as soon as it is available and the accession number will be preserved.

```

1 15827: contig of 15827 bp in length
  15828: gap of 100 bp
  15928: contig of 14557 bp in length
  30484: gap of 100 bp
  30585: gap of 100 bp
  40602: contig of 10018 bp in length
  40703: gap of 100 bp
  51912: contig of 11210 bp in length
  52013: gap of 100 bp
  60022: contig of 8010 bp in length
  60123: gap of 100 bp
  70438: contig of 10316 bp in length
  70539: gap of 100 bp
  70539: contig of 7407 bp in length
  77946: gap of 100 bp
  78046: gap of 100 bp
  85717: contig of 7672 bp in length
  85818: gap of 100 bp
  92399: contig of 6582 bp in length
  92400: gap of 100 bp
  92500: contig of 5716 bp in length
  98216: gap of 100 bp
  98316: contig of 5833 bp in length
  104148: gap of 100 bp
  104249: contig of 4918 bp in length
  109166: gap of 100 bp
  109266: gap of 100 bp
  109267: contig of 3191 bp in length
  112457: gap of 100 bp
  112458: contig of 2519 bp in length
  112558: gap of 100 bp
  115077: contig of 1887 bp in length
  115177: gap of 100 bp
  117063: contig of 1489 bp in length
  117164: contig of 1673 bp in length
  118652: gap of 100 bp
  118753: contig of 1673 bp in length
  120426: gap of 100 bp
  120526: contig of 1318 bp in length
  121844: gap of 100 bp
  121944: contig of 1249 bp in length.

```

Location/Qualifiers

```

1. .123192
  /organism="Homo sapiens"
  /mol_type="genomic DNA"
  /db_xref="taxon:9606"

```

FEATURES
source

```

/chromosome="11"
/map="11q13"
/clone="CMB9-105N5"
1. .15827
  /note="assembly_fragment"
  15928. .30484
  /note="assembly_fragment"
  30585. .40602
  /note="assembly_fragment"
  40703. .51912
  /note="assembly_fragment"
  52013. .60022
  /note="assembly_fragment"
  60123. .70438
  /note="assembly_fragment"
  70539. .77945
  /note="assembly_fragment"
  78046. .85717
  /note="assembly_fragment"
  85818. .92399
  /note="assembly_fragment"
  92500. .98215
  /note="assembly_fragment"
  98316. .104148
  /note="assembly_fragment"
  104249. .109166
  /note="assembly_fragment"
  109267. .112457
  /note="assembly_fragment"
  112558. .115076
  /note="assembly_fragment"
  115177. .117063
  /note="assembly_fragment"
  117164. .118652
  /note="assembly_fragment"
  118753. .120425
  /note="assembly_fragment"
  120526. .121843
  /note="assembly_fragment"
  121944. .123192
  /note="assembly_fragment"

ORIGIN
Query Match      18.4%; Score 460.6; DB 14; Length 123192;
Best Local Similarity 59.8%; Pred. No. 1.9e-82;
Matches 882; Conservative 1; Mismatches 575; Indels 17; Gaps 6;

QY 598 TTTGGGTTTTTTTTTTTGTGTTTGTAGAGACAGAGGTTCTGCTC-TGTACCCAGGATGA 656
Db 63120 TTTTTTTTTTTTTTTTTTTTTTTTTTTTTTGTAGACAGAGTTTCTCTCTGTTGCCAGGCTGA 63179

QY 657 GCACAGTGTGCAACCATAGGTCACTGAGCCTCAACCTCTCGAGCTCAAGGATCTGCT 716
Db 63180 GTGCAATGGCAATCTCGGTTTCAACCAACCTCCGCTCCCGGGTTCAAGGGTTTCTCC 63239

QY 717 GACCTCAGCTCCCAAGTAGCTGGGACTACGAGCGTGACACACACGCTGGCTAATTA 776
Db 63240 TGCCTCAGCTCCCGAGTAGCTGAGATTACGATGCGCCACCGCCAGCTGATTTT 63299

QY 777 AAAAAATTTTTTTGTAGAGACTGGGTTCTTACTAGTTGGCCAGGCTTGTCTTAACTCCTG 836
Db 63300 GTA---TTATTAGTAGAGAGCGGGTTTCTCCATGTTGGTTCAGGCTGTTCTCGAGCTCTG 63356

QY 837 GCTTCAGCAATCTCTCTACCTTGGCATCCCAAGTGTGGGATTAACAGGGTGAGCCAC 896
Db 63357 ACCTCAGGTGATCCACCCACCTCGGCTCCCAAGTGTGGGATTAACAGGGTGAGCCAT 63416

QY 897 CATGTGCGGCTACTTATTTTTCATATTCATCTTTTCAATAGATGAATGAATCCACAGA 956
Db 63417 CACGCCCGGCTCACATATCTTATCTTAATGTTTTTTCATGTTTGAAGTCTCAGTA 63476

QY 957 ACAGGAGTACTGCCTATTTTCTTCTTTTCTTTTGTAGACAGAGTCTCACTTCATCACC 1016

```

Db 63477 ATCATCTTTGATTCTTTTTTTTTTTGAGACAGTCTTGCTCTGTCCACCGAGCTGGAGTGCA 63536

Qy 1017 TCAACCTCGGTTGAGTCACTGCAACCTCTGCCCTCCGGGTTCAAGGATTCCTCGCCT 1076

Db 63537 ATGGCGCAATCTCGGCTCACTGCAAACTCTGCCCTCGAGGTTACGCGATTCCTCGCCT 63596

Qy 1077 AAGCCTCCTGAGTCTGGAAATCAAGCGTGACACCACTGCTTGCTGCTAATTTTTT -G 1134

Db 63597 CAGCCTCAAGATGAGTGGGATTAAGAGTGGCGGCCACACGCTGCTGCTACTTTTTTGT 63656

Qy 1135 TATTTTTAGCAGATGGGGTTTTTACCATGTGTGCCAGGCTGGTCTCAAACTCCTGACCT 1194

Db 63657 ATTTTTAGTAGAGACAGGGTTTTCAACATGTGTGCTCAGGCTGGTCTCGAACTCCTGACCT 63716

Qy 1195 CAAGTATCTGCCCTGCTCAGTCTCCCAAGTGTGGAATTAAGGCTGAGTCACTGTG 1254

Db 63717 CAGGTGATCGCCCTGCTCAGCCTTCCAAAGTGTGGGATTAACAGGATGAGCCACGCG 63776

Qy 1255 CCTGGCGGATTTACTGTCTATTTCTTTATTGCTATATCCCGAGATCTAGAGCTGCTG 1314

Db 63777 CCGATGCCCTTTGATTTCTTTGAAACAAATGATATATAATTAATGCTAAACGT-TA 63835

Qy 1315 ACATATAGTAGTGCTCAATAAATAATGATGATGCACAGCCTAGATATAAACTTTCTT 1374

Db 63836 TCATTTCTGTGATCATAAGSCTCTGACAACATTTTCTCTAGATATTGTTGTCATTATG 63895

Qy 1375 TTCTCTTTTTTAAACAACTTTGCAACTTTGCGAATAATAACAATCTTGCAATCTGCT 1434

Db 63896 ATTATTATTCACATAAAGCAAAATAATATGGGTACTACAAATTTTGTGCTATTATT 63955

Qy 1435 TTTTCA-----CTTATCACCTGTGTATGACTTTTTCATATTTGCTCAAACTTTAT 1485

Db 63956 ATATCAATANAAGATGATTAAGCAGAGAAACTAAATGTTATTTGGAAGGTATCTGAA 64015

Qy 1486 TGTACTGTTTTTTCATTTGTTACTATTTTATAGTCACTGAATAATATGCTTAATTTGCTTA 1545

Db 64016 TGGGTGTTAGGTTTTTCCCTCTCACTTAGTTAATACATACCCAGAGATGGATTTATTGT 64075

Qy 1546 TACATCTCTGCTCCACTTTAGAGGCAAAATTTACAAATCTGATGAAGCTATGAACC 1605

Db 64076 TTCTAGAACACAGACAGGAATTAGCATCTGTTTTTCTATTTTATAGATATGACCACTGAGT 64135

Qy 1606 CTCTCCCCAGAGAATAACACACACACACACTCACACAGT-TTTTTTTTAAATGTT 1664

Db 64136 CCATGGAGCTAGAAGGAATTTGTAGTGAATTTTCCCACTCTGTCGAAGTTGGAGGC 64195

Qy 1665 TGCAACTAAGACAAGAAACCTGCAATTAAGAGATGTTTGTTCATATTAATTAATAAATAACT 1724

Db 64196 CAAAAATGTCACAATGTAATCTAGAGTCAAAATTAATTTTACTTTTTTAAATGATTTGA 64255

Qy 1725 CAGTTGGGCACAGTCACTCAAGCCTGTAAACACAGTACTTTGGAAGTCCAAGTGGGTGG 1784

Db 64256 AGGCCAGGTGGGTGGCTCAGCTATGTAATCCAAGCACTTTGGAAGGCCGAGGAGTGG 64315

Qy 1785 ATCACTTGAGTGAAGTTCGAGACCGAGCTGCTCAATATGTTGAACCTCTCTCTAC 1844

Db 64316 ATTACTTGAGTCAAGGTTTCGAGACCGAGCTGGCCACATGGTGAACACTGCTCTAC 64375

Qy 1845 TAAAAATCAAAAAATTTAGCTGGGTGTAGTGTGATGCTGCTAGTCCAGCTACTCGGGA 1904

Db 64376 TAAAAATCAAAAAATTTAGCTGGGTGTGATGCTGCTGCTGCTAATCCAGCTACTCGGGA 64435

Qy 1905 GGCTGAGCAGAGAAATTTGTTGAACCTGGAGCGCAGAGTTGAGTGGAGCGGAGATCCC 1964

Db 64436 GGCTGAGCAGAGAAATTCGTTGAACCGGAGCGGAGGTTGAGTGGAGCGGAGATTTGT 64495

Qy 1965 ACCACTGACCTCCAGCCTGGCGCACACAGCAGACTCTATCTCAAAAAATAAATAAATA 2024

Db 64496 GTCACTGACCTCCAGCCTGGCGCACACAGCAGAGATCTGCTCAAAAAATAAATAAATA 64555

Qy 2025 AAATAAGGATCGGAGAGAAACAAACCTAATAAGA 2059

Db 64556 AAAAATATGAAAAATAATAAATGATGTTAAAGATA 64590

RESULT 9

AC022408

LOCUS

DEFINITION

AC022408

VERSION

KEYWORDS

SOURCE

ORGANISM

REFERENCE

AUTHORS

TITLE

JOURNAL

REFERENCE

AUTHORS

TITLE

JOURNAL

REFERENCE

AUTHORS

TITLE

JOURNAL

REFERENCE

AUTHORS

TITLE

JOURNAL

REFERENCE

AUTHORS

TITLE

JOURNAL

REFERENCE

AUTHORS

TITLE

JOURNAL

COMMENT

FEATURES

source

misc_feature

misc_feature

unSURE

misc_feature

ORIGIN

AC022408

Homo sapiens chromosome 15 clone RP11-118J16 map 15q15, complete sequence.

AC022408

GI:15217198

HTG.

Homo sapiens (human)

Homo sapiens

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini;

Hominidae; Homo.

1 (bases 1 to 150266)

Rowen, L., Madan, A., Qin, S., Baradarani, L., Birditt, B., Bloom, S.,

Burke, J., Dors, M., Fleetwood, P., Kaur, A., Madan, A., Nesbitt, R.,

Pate, D. and Hood, L.

Sequencing of human chromosome 15 D15S146-D15S117 region

Unpublished

2 (bases 1 to 150266)

Rowen, L., Madan, A., Qin, S., Baradarani, L., Birditt, B., Bloom, S.,

Bloom, S., Dors, M., Dickhoff, R., Fleetwood, P., Harrison, G.,

James, R., Kaur, A., Madan, A., Owen, M.P., Ratcliffe, A., Shaffer, T.

and Hood, L.

Direct Submission

Submitted (03-FEB-2000) Multimegabase Sequencing Center, University

of Washington, PO BOX 357730, Seattle, WA 98195, USA

3 (bases 1 to 150266)

Rowen, L., Madan, A., Qin, S., Baradarani, L., Birditt, B., Bloom, S.,

Burke, J., Dors, M., Fleetwood, P., Kaur, A., Madan, A., Nesbitt, R.,

Pate, D. and Hood, L.

Direct Submission

Submitted (21-AUG-2001) Multimegabase Sequencing Center, Institute

for Systems Biology, 4225 Roosevelt Way NE, Suite 200, Seattle, WA

98105, USA

On Aug 21, 2001 this sequence version replaced gi:11225361.

----- Genome Center Sequencing Center

Center code: UWMSC

Web site: http://chroma.mbt.washington.edu/mag_www

Contact: leerowen@systemsbiology.org

----- Summary Statistics

Sequencing vector: pUC18, 108752

Chemistry: Dye-terminator Big Dye; 90% of reads

Assembly program: Phrap; version 0.990399

Note: data from AC021753 [Drafting center UWMSC] and AC012652

[Drafting center UWMSC] and 58017 [Drafting center WUGSC] were

added for finishing.

Location/Qualifiers

1. .150266

/organism="Homo sapiens"

/mol_type="genomic DNA"

/db_xref="taxon:9606"

/chromosome="15"

/map="15q15"

/clone="RP11-118J16"

/clone_lib="RP11-118J16"

/notes="This clone overlaps RP11-46M12 AC012652 and

RP11-129112 AC021753 and RP11-58017 AC068981. Data from

overlapping BACs were combined and the consensus sequence

determined from RP11-118J16."

1. .150266

/notes="overlap with RP11-58017 AC068981."

1. .100397

/notes="overlap with RP11-129112 AC021753."

49355-49359

/notes="low quality data."

100628..150266

/notes="overlap with RP11-46M12 AC012652."

ORIGIN

Query Match 18.4%; Score 460.4; DB 8; Length 150266;
Best Local Similarity 62.0%; Pred. No. 2e-82;
Matches 889; Conservative 1; Mismatches 487; Indels 57; Gaps 8;

Qy 600 TGGGTTTTTTTGTGTTAGAGACAGGGTCTGCTGTCACCCAGGAGATGACGA 659
Db TGTGTATTTTGTGTTGTTAGAGACAGGGTCTGCTGTCACCCAGGAGATGACGA 659
Qy 660 CAGTGGTGCAACACATAGTCACTGACGCTCAACCTCTGAGCTCAAGGATCTGCTGAC 719
Db CAGTGGTCAATCTGAGTCACTGCAATCTGCTCTCTGAGCTCAAGGATCTCTCTAC 719
Qy 720 CTCAGCCTCCCAAGTAGCTGGGACTACGAGCGTGCACACACACGCTGGCTCAATTAATA 779
Db CTCAGCATCCCAAGTAGCTGGGACTACGAGTGCACATCACATGCTTAGCTAATTTTGT 7317
Qy 780 AATTTTTTTGTAGAGATGGGTCTTAACGTTGGCCAGGCTTGTCTTAACCTCTGGCT 839
Db A--TTTTTCTGTAGAGACAGGGTCTTGGCATGTTGCCATGCTGATTTTCAAACTCTGAGC 52375
Qy 840 TCAAGCAATCTCTCTACCTTGGCATCCAAAGTCTGGGATTAAGGCTGAGCCACCAT 899
Db TCAACCATCCGCTT--CTCAGCCTCCCAAGTCTGGGATTAAGGCTGAGCCACCAA 52434
Qy 900 GTGCGGCTACTTATTTCTTTACATTCATCTTTCCAAATAGATGTAAGATCCACAGAAC 959
Db GCCTAACCTATTAATGTGAATTTGTATGTTGGCATTAATATCCCTGACGATTTGT 52494
Qy 960 GGGATTAAGTCTATTTCTTTCTTTCTTTTGTAGACAGAGTCTCACTTCATCACTCA 1019
Db TTTGTT--TTGTTTTTTGTAGACAGAGTTTGTCTTGTGCACTCCACTGGAGTGCAGTG 52553
Qy 1020 ACCTCCGTTGAGTCACTGCACTCTGCTCCGAGTCAAGGATCTCTGCTGCTTAAG 1079
Db GCTCAATCTCGGCTCACTGCACTCTGCTCCGAGTCAAGGATCTCTGCTGCTTAAG 52613
Qy 1080 CCTCCTGAGTGTGGAATTAAGCGTGCACACCAATGCTGCTCAATTTTGTATTT 1139
Db CTCTGAGTGTGGAATTAAGCGTGCACACCAATGCTGCTCAATTTTGTATTT 52673
Qy 1140 TTAGCAGAGTGGGTTTTACATGTTGCTCCAGGCTGCTCAAACTCTGACCTCAAGT 1199
Db TTGAGTGTGGAATTAAGCGTGCACACCAATGCTGCTCAAACTCTGACCTCAAGT 52733
Qy 1200 GATCTGCTGCTGAGTCTCCAAAGTCTGGAATTAAGCGTGTGCTGCTGCTGCTG 1259
Db GATCCACTGCTGAGTCTCCAAAGTCTGGAATTAAGCGTGTGCTGCTGCTGCTGCTG 52793
Qy 1260 CCGATTAAGTCTGATTTCTTTTATTTGCTATATCCAGATCTAGAGAGTCTGACATA 1319
Db CCAACAACTTTTATTAAGTGAACCTCTAACACACAGCTTAATTAATGTTT 52845
Qy 1320 TAGTAGGTGCTCAATAAATGAATGCAATGCAAGCTAGATATAAATTTCTTTTCT 1379
Db TTTTGTGTTTGTAGACAGAGGCTGCTGCTCTCCACCGGCTGAG 52891
Qy 1380 TTTTGTGCTGCTGAGTCTGAGTCTGAGTCTGAGTCTGAGTCTGAGTCTGAGTCTGAG 1439
Db TGGCTGGCAATCATAGTCTACTAGGAGCTCA--GTGA 52930
Qy 1440 ACTTATCACTGTTGATGCTTTTCTATATGCTCAAACTTTATTTGTTACTGTTTTT 1499
Db TCCTCTGCTCTGAGTCTGAGTCTGAGTCTGAGTCTGAGTCTGAGTCTGAGTCTGAG 52990
Qy 1500 CATTTGTTACTTATTTAGTCACTGAATAATATGCTTAATTTGCTTATATCACTCTCTGCT 1559
Db CTCTCGATTTGTTGCTTTGCTTTAATCAATTTCCACCTTGTGTTGTTGTTGTTT 53050
Qy 1560 CCATTTAGAGGCCAAATTTAATCTGATGAAGCTATGAACCTCTCCCGCAGAA 1619
Db ATTAGTCATA-----AATTTTCTAATCAATTTGTGCTCACTACATAATTTTCTCTATCA 53105

Qy 1620 ATACACACACACACACACTCACACACAGTTTTTTTTTAATGTTGCACTAAGACAAG 1679
Db CAATTTACACAGCTTTCAAAAATTTGATTGTTTCTACTGAAATAAAACGTAGAGGAAAG 53165
Qy 1680 AAACCTTGCA--TTAGAGGATGTTGTTCAATATTAATAATAAATACTAGTTGGGCACA 1736
Db AACACAGTAGCTAAATGATGAATCAATCTTTAAACCAAAATTCAGCGGGCAGC 53225
Qy 1737 GTCACTCAAGCCTGTAACACACAGTACTTTTGGAGTCCAGGTGGGTGATCACTTCAGGT 1796
Db GTGGCTCAGCCTGTAATCTCAGCACTTTGGGAGGCCAAGGTAGGTGGATCAC--GAGGT 53283
Qy 1797 GAGAAGTTGAGACACAGCCTGCTCAATATGTTGAAACCTCTATCTTACTAAAAATACAA 1856
Db CAGGGCTCGAGACACAGCCTGACCAATGTTGAAACCTGCTCTTACTAAAAATACAA 53343
Qy 1857 AATAGCTGGTGTGATGATGCTGTAGTCCAGTCTCTCGGAGGCTCAGGCAAG 1916
Db AATAGCTGGTGTGATGCTGTAGTCCAGTCTCTCGGAGGCTCAGGCAAG 53403
Qy 1917 AGAATTTGCTTGAACCTGGGAGGAGGAGTTCAGTCCAGTCCAGGAGTCCAGGCTC 1976
Db AGAATTTGCTTGAACCTGGGAGGAGGAGTTCAGTCCAGTCCAGGAGTCCAGGCTC 53463
Qy 1977 CAGCCTGGGCGACACACAGCAGACTCTATCTCAAAAAATAAATAAATAAATAA 2030
Db CAGCCTGGGCGACACAGCAGACTCTATCTCAAAAAATAAATAAATAAATAAATAA 53517

RESULT 10
AC021753

LOCUS Homo sapiens chromosome 15 clone RP11-129I12 map 15q15, complete sequence.
DEFINITION AC021753
ACCESSION AC021753
VERSION AC021753.7 GI:16507213
KEYWORDS HTG.
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 167996)
AUTHORS Rowen, L., Madan, A., Qin, S., Baradarani, L., Birditt, B., Bloom, S., Burke, J., Dors, M., Fleetwood, P., Kaur, A., Madan, A., Nesbitt, R., Pate, D. and Hood, L.
TITLE Sequencing of human chromosome 15 D15S146-D15S117 region
REFERENCE 2 (bases 1 to 167996)
AUTHORS Rowen, L., Madan, A., Qin, S., Abbasi, N., Baradarani, L., Birditt, B., Bloom, S., Dors, M., Dickhoff, R., Fleetwood, P., Harrison, G., James, R., Kaur, A., Madan, A., Owen, M.P., Ratcliffe, A., Shaffer, T. and Hood, L.
TITLE Direct Submission
JOURNAL Submitted (20-JAN-2000) Multimegabase Sequencing Center, University of Washington, PO BOX 357730, Seattle, WA 98195, USA
REFERENCE 3 (bases 1 to 167996)
AUTHORS Rowen, L., Madan, A., Qin, S., Baradarani, L., Birditt, B., Bloom, S., Burke, J., Dors, M., Fleetwood, P., Kaur, A., Madan, A., Nesbitt, R., Pate, D. and Hood, L.
TITLE Direct Submission
JOURNAL Submitted (29-OCT-2001) Multimegabase Sequencing Center, Institute for Systems Biology, 4225 Roosevelt Way NE, Suite 200, Seattle, WA 98105, USA
COMMENT On Oct 29, 2001 this sequence version replaced gi:11136698.
----- Genome Center
Center: Multimegabase Sequencing Center
Web site: http://chroma.mbt.washington.edu/msg_www
Contact: leerowen@systemsbiology.org
----- Summary Statistics
Sequencing vector: pUC18; 108752
Chemistry: Dye-terminator Big Dye; 90% of reads

Chemistry: Dye-primer Big Dye; 10% of reads
Assembly program: Phrap; version 0.990399
Note: data from AC022408 [Drafting center UHMSC] and AC020661
were added for finishing.
Location/Qualifiers
source
1. .167996
/organism="Homo sapiens"
/mol_type="genomic DNA"
/db_xref="taxon:9606"
/chromosome="15"
/map="15q15"
/clone="RP11-129112"
/clone_lib="RP11-129112"
/notes="This clone overlaps RP11-118J16 AC022408 and
RP11-540011 AC020661. Data from overlapping BACs were
combined and the consensus sequence determined from
RP11-129112 to the extent possible."
misc_feature
1. .63402
/notes="overlap with RP11-540011 AC020661."
misc_feature
67596.. .167996
/notes="overlap with RP11-118J16 AC022408."
ORIGIN
Query Match 18.4%; Score 459.4; DB 8; Length 167996;
Best Local Similarity 62.0%; Pred. No. 3.1e-82;
Matches 888; Conservative 1; Mismatches 487; Indels 57; Gaps 8;
Qy 600 TGGGTTTTTTTGTGTTAGAGACAGGGTCTGTGTCACCCAGGCAATGACGA 659
Db 119755 TGTGTAATTTTGTGTTGTTAGCCAGCTCGGGCTCTGTGCGCCAGGCTGGAGTG 119814
Qy 660 CAGTGGTGCAACATAGTCACTGACGCTCAACCTCTGAGCTCAAGGATCTGTGAC 719
Db 119815 CAGTGGTCTGATCTCGGCTCACTGCAATCTCTGCTCTGGGCTCAAGCCATCTCTCAC 119874
Qy 720 CTGAGCTCCCAAGTAGCTGGAGCTAGCAGCGTGCACACACACGCTCGGCTTAATTAATA 779
Db 119875 CTCAGATCCCAAGTAGCTGGAGCTACAGTGCACATCCATGCTTAGCTAATTTTGT 119934
Qy 780 AATTTTTTTGAGAGCTGGTCTTACTAGCTGGCAGGCTGTGCTTAACTCTGGCT 839
Db 119935 A--TTTCTGTAGAGACAGGGTTTTGGCATGTTGCCCATGCTGATTTCAAACTCCTGAGC 119992
Qy 840 TCAGCAATCTCTACTCTTGGATCCCAAGTCTGGATACAGGGGTGAGCCACAT 899
Db 119993 TCAAACGATCGGCT-GCTCAGCTCTCCAAAGTCCGGGATACAGGTGTGAGCCCAA 120051
Qy 900 GTGGGGTACTTATTTCTTACATTCATCTTTCCATAGAAATGTAAGATCCACAGAAC 959
Db 120052 GCCTAACTTATTAATGTGAATTTGTATGTTGGCATATTAATCCCTAGCCAGTTTGT 120111
Qy 960 GGGATTAAGTCTATTTCTTCTTTCTTTTGTGAGACAGATCTCACTCATCACTCA 1019
Db 120112 TTGCTT-TTGTGTTTGTGAGACAGATTTTGTCTTGTGCACTCCACTGGAGTGCAGTG 120170
Qy 1020 ACCTCCGTTAGCTCACTGCAACTCTGCTCCCGGCTCAAGGATTCCTGCTCAAG 1079
Db 120171 GCTCAATCTGGCTCACTGCAACTCTGCTCCCGGCTCCCGGGTTCAGGGGATTCCTCCCTCAG 120230
Qy 1080 CCTCTGAGTAGCTGGATTAACAAGCGTGCACCACTGCTTTGGCTAATTTTGTATTT 1139
Db 120231 CCTCTGAGTAGCTGGATTAACAGTGGCGCCACCGCCAGCTAATTTTGTATTT 120290
Qy 1140 TTAGAGAGATGGGGTTTTACATGTTGCCAGGCTGGTCTCAAACTCTGACCTCAAGT 1199
Db 120291 TTGGTAGAGTCAGAGTTTAAACATGTTGGCCAGGCTGGTTTCGAACCTCTGACCTCAGGT 120350
Qy 1200 GATCTGCTGCTCAGTCTCCCAAGTCTCGAATATAGCGGTGAGTCACTGTCGCTGG 1259
Db 120351 GATCACTGCTCAGTCTCCCAAGTCTCGAATATAGCGGTGAGTCACTGTCGCTGG 120410
Qy 1260 CCGATTACTGCTATTTCTTTTATTTGCTATATCCCCAGATCTAGAGCAGTGTCTGACATA 1319

Db 120411 CCAACAAACTTTCAATTAAGTGAACCTCTAAACACACAGCTTATAATTAGTTT----- 120462
Qy 1320 TAGTAGGTCTCAATAAATAAATTAATGAATGACAGCCCTAGATATAAACTTTCTTTCT 1379
Db 120463 -----TGTTTGTGTTTTTAGAGACAGGGTCTCGCTCTCCACCCAGGCTTGAG 120508
Qy 1380 TTTTAAACAACTTTGACAACTTTGACAGATAAATAACAATCTTGCAATTTCTGCTTTTTC 1439
Db 120509 TGGCTGGGCAATATATAGCTCACTAGGACTCA-----GTGA 120547
Qy 1440 ACTTATCACTGTTTATGACTTTTTCATATGCTCAAACTTTATGTTAGTGTGTTTTT 1499
Db 120548 TCCTCTGCTCTCTGAGTAACCTGGGACTACAGGCTTGAGACACCATGCCCAGAGAATAG 120607
Qy 1500 CATGTTACTATTTTACTCACTGAATAATATGCTTAATTTGCTTATACATCTCTGCT 1559
Db 120608 CTCGATTTGTTGCTTTGCTTTAATCAATTTCCACCTTGATGTACTCAAAATGTTTT 120667
Qy 1560 CCACCTTTAGAGGCCAAATTTTCAAAATCTGATGAAAGCTATGAACCTCTCTCCCAGAGAA 1619
Db 120668 ATTAGTCATA-----ACTTTTCAATCACTTTGCTCACTACATAATTTTCTTATCA 120722
Qy 1620 ATACACACACACACACACTCACACAGTTTTTTTTTAAGTTTGTGCAACTAAGACAAG 1679
Db 120723 CAATTTACAGACTTTCAAAAATTTGATTTGCTTTTCTACTGAAATAAAACGTAGAGGAAAG 120782
Qy 1680 AAACCTGCA---TTAGAGGATGTTGTTTCAATATTAATTAATAATACTCACTGGGCACA 1736
Db 120783 AACACAGTAAGCTAATAATGATGAATTCATCCCTTTAAACCAAAATTCAGCCGGGACG 120842
Qy 1737 GTGACTCAAGCCTGTAAACACAGTACTTTGGAAGTCCAAGGTGGGTGGATCACTTGAAGT 1796
Db 120843 GTGGCTCAAGCCTGTAAATCTCAGCACTTTGGGAGGCCAAGTAGTGGATCAC--GAGGT 120900
Qy 1797 GAGAAGTTCGAGACCAAGCTGTGCAATATGGTGAACCCCTATCTCTACTAAATAACAAA 1856
Db 120901 CAGGGCTCGAGACCAAGCTGACCAACATGGTGAACCCCGTCTCTACTAAATAACAAA 120960
Qy 1857 AATTAGTGGTGTAGTGTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 1916
Db 120961 AATTAGTGGTGTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 121020
Qy 1917 AGAATGCTTGAACCTGGGAGGAGAGGTTGCAAGTGGAGGAGAGTCCACACCTGCACTC 1976
Db 121021 AGAATGCTTGAACCTGGGAGGAGGAGTGGAGTGGAGTGGAGTGGAGTGGAGTGGAGT 121080
Qy 1977 CAGCCTGGGCGACACAGCAGACTCTATCTCAAAAAATAATAATAATAATAATAATAATA 2029
Db 121081 CAGCCTGGGCGACAGACGAGACTCGCTCTCAAAAAACAAAAAATAATAATAATAATAATA 121133
RESULT 11
AC003689
LOCUS
DEFINITION
Homo sapiens Chromosome 11q12.2 PAC clone pDU1081b4 containing
human mRNA for T-cell glycoprotein CD6, complete sequence.
AC003689
AC003689.1 GI:3900834
VERSION
HTG.
KEYWORDS
SOURCE
Homo sapiens
ORGANISM
Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Euthera; Euarchontoglires; Primates; Catarrhini;
Hominiidae; Homo.
REFERENCE
1 (bases 1 to 137693)
AUTHORS
Evans, G.A., Athanasiou, M., Aguayo, P., Armstrong, D., Basit, M.,
Buetner, J., Bumeister, R., Card, P., deSailboat, F., Dunn, J.,
English, C., Ethridge, S., Garner, H.R., Gee, V., Gordon, M., Gotway, G.,
Grant, O., Hahner, L., Joslin, J., Lewis, E., Loo, H., Loo, K.N.,
Major, T., McFarland, J., Newcom, J., Osborne-Lawrence, S.,
Schageman, J., Schultz, R.A., Stimson, S., Syed, M. and Ward, T.
HTGS Submission
TITLE
Unpublished
JOURNAL

repeat_region	complement (63993). .64088)
repeat_region	/rpt_family="MIR". .66870)
repeat_region	/rpt_family="Alu". .67416. .67584
repeat_region	/rpt_family="MER42". .67594. .67887
repeat_region	/rpt_family="Alu". .68119. .68182
repeat_region	/rpt_family="MER42". .69519)
repeat_region	/rpt_family="Alu". .69821)
repeat_region	/rpt_family="Alu". .70676)
repeat_region	/rpt_family="Alu". .70906)
repeat_region	/rpt_family="MER42". .71247. .71537
repeat_region	/rpt_family="Alu". .71655)
repeat_region	/rpt_family="MER42". .72778. .73053
repeat_region	/rpt_family="Alu". .75330)
repeat_region	/rpt_family="Alu". .76879)
repeat_region	/rpt_family="Alu". .77232)
repeat_region	/rpt_family="Alu". .77719. .77994
repeat_region	/rpt_family="Alu". .78902. .79179
repeat_region	/rpt_family="Alu". .79179

Query Match

Best Local Similarity 18.3%; Score 457.4; DB 8; Length 137693;

Matches 880; Conservative 1; Mismatches 577; Indels 17; Gaps 6;

Qy	598	TTTTGGGTTTTTTTCTGTTGTTTGTAGACACAGGCTGTGCTC-TGTCAACCCAGGCATGA	656
Db	76572	TTTTCTTTCTTTCTTTTTTTTTTTTTTTTTTTTGTAGACACAGGTTCTCTCTTTGTGCCAGGCTGGA	76631
Qy	657	GCACAGTGGTGCAACCATAGTGCACATGCGAGCTCAACCTCTCGAGCTCAAGGGATCTGCT	716
Db	76632	GTGCATGGCACAAATCTCGGTTACACAACTCTCCGCTCCCGGGTTCAAGGTTTCTCC	76691
Qy	717	GACCTCAGCCTCCCAAGTAGCTGGGACTACGACGGTGACACCAACCGCTGGCTAATTAA	776
Db	76692	TGCTCAGCCTCCGAGTAGCTGAGATTACAGGCATGGCCACCAACCGCCAGCTGATTTT	76751
Qy	777	AAAAATTTTTTTGTAGACATGGGCTTACTAGTTGGCAGGCTTGCTTAACTCCTG	836
Db	76752	GTA---TTATTAGTAGACAGCGGGTTTCTCCATGTTGGTCAGGCTGGTCTCGAGCTCCTG	76808
Qy	837	GCCTTCAGCAATCCTCTACTTTGGCATCCCAAAGTCTGGGATTACAGGGGTAGCCAC	896
Db	76809	ACCTCAGGTGATCCACCCACTCGGCTCCCAAAGTCTGGGATTATAGGTGTAGCCAT	76868
Qy	897	CATGTGCGGCTACTTATTTCTTTACATTCCTCTTTTCCAAATAGAATGTAAAGTCCACAGA	956
Db	76869	CACGCCGGCTCACATATTTCTATCTTAATATGTTTTTCATGCTTTGAAGTCTCAGTA	76928
Qy	957	ACAGGATTAAGTGGCTATTTTCTTCTCTTTTGTGAGACAGAGTCTCACTTATCACC	1016
Db	76929	ATCATCTTTTGAATCTTTTTTTTTTTTGTAGACAGTCTTGCTCTGCTCCACCAAGGCTGGAGTGA	76998
Qy	1017	TCAACCTCCGTTTCAGTCACTGCAACTCTGCTCCCGGGTTCAAGYGATTTCTCTGCT	1076
Db	76989	ATGGCGAATCTCGGCTCACTGCAAACTCTGCTCCGAGGTTTCAACGCAATTTCTCTGCT	77048
Qy	1077	AAGCCTCTGAGTAGCTGGAAATTAAGCGGTGCACCAACCATGCTGGCTAAATTTTTTT--G	1134

ORGANISM Papio anubis
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini;
 Cercopithecoidea; Cercopithecinae; Papio.
 REFERENCE 1 (bases 1 to 195986)
 AUTHORS Akhter, N., Antonellis, A., Ayele, K., Beckstrom-Sternberg, S.M.,
 Benjamin, B., Blakeley, R.W., Bouffard, G.G., Brinkley, C., Brooks, S.,
 Cariaga, K., Coleman, B., Engle, J., Granite, S., Guan, X., Gupta, J.,
 Haghighi, P., Han, J., Hansen, N., Ho, S.-L., Idol, J.R., Karlins, E.,
 Lario, P., Lee-Lin, S.-Q., Legaspi, R., Maduro, Q.L., Maduro, V.B.,
 Margulies, E.H., Masiello, C., Maskeri, B., McDowell, J.,
 Peguirigan, C., Pearson, R., Portnoy, M.E., Prasad, A.,
 Reddix-Dugue, N., Schandler, K., Schueler, M.G., Sison, C.,
 Stantripop, S., Thomas, J.W., Thomas, P.J., Touchman, J.W., Vogt, J.L.,
 Wetherby, K.D., Wiggins, L., Young, A. and Green, E.D.
 TITLE NISC Comparative Sequencing Initiative
 JOURNAL Unpublished
 REFERENCE 2 (bases 1 to 195986)
 AUTHORS Green, E.D.
 TITLE Direct Submission
 JOURNAL Submitted (03-APR-2002) NIH Intramural Sequencing Center, 8717
 Grovemont Circle, Gaithersburg, MD 20877, USA
 REFERENCE 3 (bases 1 to 195986)
 AUTHORS Green, E.D.
 TITLE Direct Submission
 JOURNAL Submitted (21-DEC-2002) NIH Intramural Sequencing Center, 8717
 Grovemont Circle, Gaithersburg, MD 20877, USA
 COMMENT On Dec 21, 2002 this sequence version replaced gi:19909407.
 ----- Genome Center
 Center: NIH Intramural Sequencing Center
 Center code: NISC
 Web site: http://www.nisc.nih.gov
 Contact: nisc.zoem@nigri.nih.gov
 ----- Project Information
 Center project name: cqw
 Center clone name: 001C24

The sequence data in this record represents an 'enhanced' version of a Phase 2 submission. Specifically, the indicated order and orientation of each sequence contig has been established using one or more of the following: read-pair data from individual subclones, overlaps with neighboring clones, alignment with available reference sequence (e.g., human), and/or confirmation by PCR testing. In addition, the sequence assembly is based on at least 8X average coverage in Q20 bases and has been reviewed to rule out gross misassemblies, the low-quality ends of sequence contigs have been trimmed away, and each base is associated with a Phrap-derived quality score.

----- Summary Statistics
 Sequencing vector: plasmid; n/a; 100% of reads
 Chemistry: Dye-terminator Big Dye; 100% of reads
 Assembly program: Phrap; version 0.990319
 Consensus quality: 195122 bases at least Q40
 Consensus quality: 195370 bases at least Q30
 Consensus quality: 195594 bases at least Q20
 Insert size: 186000; agarose-fp
 Insert size: 195686; sum-of-contigs
 Quality coverage: 9.19x in Q20 bases; agarose-fp
 Quality coverage: 8.74x in Q20 bases; sum-of-contigs

* NOTE: This is a 'working draft' sequence. It currently consists of 4 contigs. Gaps between the contigs are represented as runs of N. The order of the pieces is believed to be correct as given, however the sizes of the gaps between them are based on estimates that have been provided by the submitter.
 * This sequence will be replaced
 * by the finished sequence as soon as it is available and
 * the accession number will be preserved.
 * 1 76152: contig of 76152 bp in length
 * 76153 76252: gap of unknown length
 * 76253 126668: contig of 50416 bp in length

* 126669 126768: gap of unknown length
 * 126769 177394: contig of 50626 bp in length
 * 177395 177495: gap of unknown length
 * 177495 195986: contig of 18492 bp in length.
 FEATURES
 source
 1. .195986
 /organism="Papio anubis"
 /mol_type="genomic DNA"
 /db_xref="taxon:9555"
 /clone_lib="RP41-1C24"
 /clone_lib="RP41"
 1. .76152
 /notes="assembly_fragment"
 clone_end:T7
 vector_side:left"
 misc_feature
 1. .54626
 /notes="clone overlaps with GenBank Accession Number AC116934 clone RP41-206F5 (center project name cqw)"
 76153. .76252
 /estimated_length=unknown
 misc_feature
 76253. .126668
 /notes="assembly_fragment"
 126669. .126768
 /estimated_length=unknown
 misc_feature
 126769. .177394
 /notes="assembly_fragment"
 177395. .177494
 /estimated_length=unknown
 misc_feature
 177495. .195986
 /notes="assembly_fragment"
 clone_end:SP6
 vector_side:right"
 ORIGIN
 Query Match 18.1%; Score 451.2; DB 14; Length 195986;
 Best Local Similarity 61.1%; Pred. No. 1.3e-80;
 Matches 891; Conservative 1; Mismatches 514; Indels 52; Gaps 8;
 QY 604 TTTTCTTTTCTGTTCTTTAGACAGAGGCTTGTCTGTCCACCCAGGATGAGCAGT 663
 Db 180502 TTTTCTTTTCTGTTCTTTAGACAGAGGCTTGTCTGTCCACCCAGGATGAGTGCAGT 180561
 QY 664 GGTGCAACCATAGGTCACTGCAGCCTCAACCTCTGAGCTCAAGGGATCTGTGACCTCA 723
 Db 180562 AGTGCATAGTGTTCCTTCATCTTCTGGAACAGTGTATCTCCACCTCA 180621
 QY 724 GCTCCCAAGTAGCTGGACTAGAGGTGCAACACAGCCTGGCTTAATTAATAAATT 783
 Db 180622 GTCTTCTGAGTAGCTGGGACTTTAGGCACATGCCCAGCTAACTTAAAAAATT 180681
 QY 784 TTTTGTAGAGACTGGGTCTTACTAGTGTGGCCAGGCTTGTCTTAAACTCTGGCTTCAA 843
 Db 180682 TTTTGAAGACACAGGCTCTGTCTGTCTGCTCAGGCTACTCTCAAAACAGCCGCTCAA 180741
 QY 844 GCAATCTCTTACTTGGCATCCAAAGTGTCTGGGATTAAGGGGTGAGCCACCATGTGC 903
 Db 180742 GTGATCTCTCGCTCATCTCCCAAGTGTCTGGATAGAGGTGTGAGCCACCATGCC 180801
 QY 904 GGCTACTATTCTTTTACATTCATCTTCCAAATAGATGTAAGATCCACAGAACAGGGA 963
 Db 180802 AGCCCATTTTTTTTAAATTTCTTAATTTCTATTTTAAATTTTAAATTTTAA 180858
 QY 964 TTACTGCCTATTTTCTTCTCTTTTGTGAGACAGTCTCACTCATCACTCAACCT 1023
 Db 180859 TTTATTTTATTTTATTTATTTATTTTATTTTATTTTATTTTATTTTATTTTATTT 180918
 QY 1024 CCGT-----TCAGCTCACTGCAACTCTGCTCCCGGGTTCAGGATTTCT 1069
 Db 180919 GAGTGCAATGGTGTGATCTCAGCTCACTGCAACTCTCGCTCCCGGGTTCAGCAACTCC 180978
 QY 1070 CTTGCCTTAAGCCTTCCTGAGTAGCTGGAATTAAGCGGTGCACCACTGCTTGGCTAATT 1129
 Db 180979 CCTGCCTCAGCCTCCTGAGTAGCTGGGATTAACAGGTGGCCTCCACTACACCTGGCTAA-T 181037

Qy 855 ACCTTGGCATCCCAAGTGTCTGGGATTAACAGGGGTGAGCCACCATGTGCGCTACTTATT 914
Db 53494 GCCTTGGCCTCCCAAGTGTCTAGGATTAACAGGCATGAGCCACTGCACCCGCGCTGATTCC 53435
Qy 915 TCTTTACATTCATCTTCTTCCAAATAGATGTAAGATCCACAGAACAGGGATTTACTGCCAT 974
Db 53434 TGAATCTCTTTCTTTCTTTT-----TTCTTTCT 53409
Qy 975 TTCTCTCTCTTTCTTTTGTAGACAGAGTCTCACTTTCATCACTCAACCTCCGTTCA--- 1030
Db 53408 TTTTCTTTTGTAGACAGAGTCTCGCTCTGTGCGCCAGGCTGGAGTGCAGTGG 53349
Qy 1031 -----GCTCACTGCAACCTCTGCTCTCCCGGGTTCAAGYGATTTCTCTGCTTAAGC 1080
Db 53348 CACGATCTTGGCTCACTGCAAGTCCACCTCTCTGGTTTCAGGCCATCTCTGCTCAGC 53289
Qy 1081 CTCCTGAGTAGCTGAATTAACAGCGTGCACACATGCTGGCTAAATTTTGTATTTT 1140
Db 53288 CTCGCGAGTAGCTGGGACTACAGGTGCCCGCCACACACCCAGCTAAATTTTGTATTT 53229
Qy 1141 TAGCAGAGATGGGGTTTTTACCATGTTGCCAGGCTGTCTCAAACTCTGACCTCAAGTG 1200
Db 53228 TAGTAGAGCGGGTTTACCGTTTACCGATGTTTACAGGATGTTCTGATCTCTGACCTC--GTG 53171
Qy 1201 ATCTGCTGCTCAGTCTCCCAAGTGTCTGGAATTTATAGGCGTGAGTCACTGTGCTTGGC 1260
Db 53170 ATCCGCTGCTCGGCTCCCAAGTGTCTGGAATTTACAGCATGAGCCACCGCGCTTGGC 53111
Qy 1261 CGATTACTGTCTATTTCTTTATTTGCTATATCCCGAGATCTAGAGAGTGTCTGACATAT 1320
Db 53110 CTTTTCTTTTATTTTATTTTGTAGACAGGGCTTC-CTCTGTCCACCGAGCTGTAATGC 53052
Qy 1321 AGTAGGTGCTCAATAAATAATGATGAATGACAGCGCTAGATATAAACTTCTTTTCTT 1380
Db 53051 ACTGGCACTATCTTGGCTCACTGCAACCTCCACTTCCAGGTTCAAGCAATATATCATGCC 52992
Qy 1381 TTTTAAACCAATCTTGACAACTTTGACAAATAAATAAATCTTGTCAATCTGCTTTTCA 1440
Db 52991 TCAGTCTCTGAGTAACTGGGATTTACAGGCGTGTGCGCTACGCCAGCTAATTTTGTGTA 52932
Qy 1441 CTTATCACTTGTATGACTTTTTTATATATGCTTCAAAAC---CTTATTTGTACTGTTTT 1497
Db 52931 TTTTATAGAGAGGAGGTTTTCCCATGTGTGGCAAGATGCTCTCCATCTCTTGACCTCG 52872
Qy 1498 TTCATCTGTACTATTTTAGTCACTGAATAATATGCTTTAATTTGCTTATACATCTCTCTG 1557
Db 52871 TGATCTCCCGCTCGGCGCTCCCAAGTGTGAGATTTACAGGCGTGAGCCACTGTGCGCG 52812
Qy 1558 CTCACCTTTAGAGGCGCAAAATTTTACAAATCTGATGAAGCTATGAACCTCTCCCCAGAG 1617
Db 52811 GCCTGATAACAGCTTTCTATGTGCCAGCTACTGCTCCAGTTTATATGATATATGAT 52752
Qy 1618 AATATACACACACACACACACTACACAGTTTTTTTTTTAAATGTTTGCACCTAAGACA 1677
Db 52751 TAAATTCATTAGCTCAITTTGATTTATATGATCTATTAATTCCTTTTATTAGTCCATTTTC 52692
Qy 1678 AGAAACCTGATTTAGAGGATTTTGTTCATTAATTAATAAATAAATCACTAGTTGGGCACAG 1737
Db 52691 ACCCTGCTGATAAGACATCTTGAGACTGAGAAAGAAAGAGTGTTTAGGCCAGGAGTGG 52632
Qy 1738 TGATCAAGCTGTAAACACAGTACTTTGGAAGTCCAAAGTGGGTGGATCACTTTGAGGTG 1797
Db 52631 TAGCTCACACTGTATCCAGACCTTTGGAGGCGGAGGTGGAAGATTTACCTGAGATC 52572
Qy 1798 AGAAGTTGAGACAGCCTCGTCAATGATGTGAAACCTTATCTTACTAAAAATACAAAA 1857
Db 52571 AGGAGTTTGAAGACACCTCGGCCAAACATGCAAAACCCCATCTCTACTAAAAATACAAAA 52512
Qy 1858 ATTAGCTGGGTGTAGTATGATCGCTGTAGTCCAGCTACTCGGGAGGCTGAGGACAGA 1917
Db 52511 ATTAGTAGGCGATGGTGGTGTACATGCTGCTGATCTCAGCTACTTTGGGAGGCTGAGGACGA 52452

Qy 1918 GAATTGCTTGAACCTGGAGCGAGAGTGTGAGTGGAGCGAGATCCACCACTGCACTCC 1977
Db 52451 GAATTGCTTGAACCTGGAGGTGGAGGTGTAGTGGAGCCAGGATCACGCCACTGCACTCC 52392
Qy 1978 AGCTGGGCGACACAGCGAGACTTCTTCAAAAAAATAAATAAATAAAGGATCG 2037
Db 52391 AGCTGGGTGACAGAAATGAGACTCCATCTTAAAAAATAAATAAAGGAAAAAGAA 52332
Qy 2038 GAGAGA 2043
Db 52331 AAAAGA 52326
RESULT 15
AC005484/c
LOCUS 131943 bp DNA linear PRI 30-SEP-2000
DEFINITION Homo sapiens PAC clone RP5-84708 from 14q24.3, complete sequence.
ACCESSION AC005484
VERSION AC005484.2 GI:5091654
KEYWORDS HTG.
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini;
Hominoidea; Homo.
REFERENCE 1 (bases 1 to 131943)
AUTHORS Sulston, J.E. and Waterston, R.
TITLE Toward a complete human genome sequence
JOURNAL Genome Res. 8 (11), 1097-1108 (1998)
PUBMED 9847074
REFERENCE 2 (bases 1 to 131943)
AUTHORS Cloud, J., Wohlmann, P., and Holmes, A.
TITLE The sequence of Homo sapiens PAC clone RP5-84708
JOURNAL Unpublished
REFERENCE 3 (bases 1 to 131943)
AUTHORS Waterston, R.H.
TITLE Direct Submission
JOURNAL Submitted (14-AUG-1998) Genome Sequencing Center, Washington
University School of Medicine, 4444 Forest Park Parkway, St. Louis,
MO 63108, USA
REFERENCE 4 (bases 1 to 131943)
AUTHORS Waterston, R.H.
TITLE Direct Submission
JOURNAL Submitted (17-JUN-1999) Genome Sequencing Center, Washington
University School of Medicine, 4444 Forest Park Parkway, St. Louis,
MO 63108, USA
REFERENCE 5 (bases 1 to 131943)
AUTHORS Waterston, R.
TITLE Direct Submission
JOURNAL Submitted (30-SEP-2000) Department of Genetics, Washington
University, 4444 Forest Park Avenue, St. Louis, Missouri 63108, USA
On Jun 17, 1999 this sequence version replaced gi:3907511.
----- Genome Center
Center: Washington University Genome Sequencing Center
Center code: WUGSC
Web site: <http://genome.wustl.edu/gsc>
Contact: sapiens@wustl.edu
----- Summary Statistics

Center project name: H_DJ0847008

NOTICE: This sequence may not represent the entire insert of this clone. It may be shorter because we only sequence overlapping clone sections once, or longer because we provide a small overlap between neighboring data submissions.

This sequence was finished as follows unless otherwise noted:
all regions were double stranded, sequenced with an alternate chemistry, or covered by high quality data (i.e., phred quality >= 30); an attempt was made to resolve all sequencing problems, such as compressions and repeats; all regions were covered by sequence from more than one subclone; and the assembly was confirmed by restriction digest.

MAPPING INFORMATION:

This clone from chromosome 14 was provided by Dr. Pieter de Jong, Roswell Park Cancer Institute, Human Genetics Department, Elm and Carlton Streets, Buffalo NY 14263-0001 USA.

SOURCE INFORMATION:

This clone was derived from human PAC library RPCI-5, prepared by Pieter de Jong and coworkers at the Roswell Park Cancer Institute (http://bacpac.med.buffalo.edu) using the method described by Ioannou et al., Nature Genetics 6:84-9 (1994). The library is from one male donor.

The clone may be obtained either from Genome Systems, Inc.

(http://www.genomesystems.com) or Research Genetics, Inc.

(http://www.resgen.com); or from Pieter de Jong.

VECTOR: pCYPAC2

NEIGHBORING SEQUENCE INFORMATION:

The clone sequenced to the left is RP5-892G5; the clone sequenced to the right is RP4-592G7. 200 bp overlap. Actual start of this clone is at base position 1 of RP5-84708; actual end is at base position 131747 of RP5-84708.

FEATURES

source

1. .131943
/organism="Homo sapiens"
/mol_type="genomic DNA"
/db_xref="taxon:9606"
/chromosome="14"
/map="14q24.3"
/clone="RP5-84708"
/clone_lib="RPCI-5"
1058. .1098
/rpt_family="AT_rich"
1762. .2225
/rpt_family="MaLR"
2305. .2587
/rpt_family="Alu"
2937. .3086
/rpt_family="L2"
3293. .3591
/rpt_family="Alu"
3651. .3916
/rpt_family="L1"
3917. .3943
/rpt_family="(TA)n"
4815. .4841
/rpt_family="AT_rich"
5153. .5244
/rpt_family="L1"
5728. .6008
/rpt_family="Alu"
6009. .6034
/rpt_family="(CAA)n"
8080. .8468
/note="match to EST R05271 (NID:g755891) ye91b03.s1"
8080. .8277
8284. .8589
/rpt_family="Alu"
8649. .9039
/note="match to EST R05328 (NID:g755948) ye91b03.r1"
8857. .9007
/rpt_family="MIR"
9062. .9152
/rpt_family="MIR"
9216. .9288
/rpt_family="MIR"
9476. .9770
/rpt_family="Alu"
9711. .10034
10004. .10114
/rpt_family="Alu"
10115. .10135
/rpt_family="(TAAA)n"
10136. .10158

repeat_region
10205. .10501
/rpt_family="Alu"
repeat_region
10513. .10819
/rpt_family="Alu"
repeat_region
10890. .11242
/rpt_family="MaLR"
repeat_region
11246. .11519
/rpt_family="Alu"
repeat_region
11567. .11996
/rpt_family="L2"
repeat_region
12006. .12113
/rpt_family="L2"
repeat_region
12123. .12146
/rpt_family="(CAAAA)n"
repeat_region
12147. .12430
/rpt_family="Alu"
repeat_region
12445. .12535
/rpt_family="L1"
repeat_region
12527. .13097
/rpt_family="L1"
repeat_region
13217. .13339
/rpt_family="Alu"
repeat_region
13389. .13994
/rpt_family="L2"
repeat_region
14138. .14158
/rpt_family="AT_rich"
repeat_region
14195. .15071
/rpt_family="L1"
repeat_region
15048. .15521
/rpt_family="L1"
repeat_region
15520. .15992
/rpt_family="L1"
repeat_region
15989. .16221
/rpt_family="L1"
repeat_region
16257. .16556
/rpt_family="Alu"
repeat_region
16569. .16963
/rpt_family="L1"
repeat_region
17016. .17039
/rpt_family="AT_rich"
repeat_region
17073. .17182
/rpt_family="MIR"
repeat_region
17252. .17452
/rpt_family="MaLR"
repeat_region
17473. .17991
/rpt_family="L2"
repeat_region
18106. .18192
/rpt_family="MER1_type"
repeat_region
18193. .18504
/rpt_family="Alu"
repeat_region
18505. .18588
/rpt_family="MER1_type"
repeat_region
18624. .18904
/rpt_family="L1"
repeat_region
19144. .19368
/rpt_family="Alu"
repeat_region
20030. .20159
/rpt_family="Alu"
repeat_region
20167. .20317
/rpt_family="Alu"
repeat_region
20318. .20617
/rpt_family="Alu"
repeat_region
20618. .20753
/rpt_family="Alu"
repeat_region
20802. .20954
/rpt_family="L1"
repeat_region
20955. .21256
/rpt_family="Alu"

Query Match 17.9%; Score 447.8; DB 8; Length 131943;
Best Local Similarity 61.3%; Pred. No. 6.8e-80;
Matches 913; Conservative 1; Mismatches 523; Indels 52; Gaps 10;

THIS PAGE BLANK (USPTO)

GenCore version 5.1.7
Copyright (c) 1993 - 2006 Bioacceleration Ltd.

OM nucleic - nucleic search, using sw model

Run on: February 9, 2006, 13:28:05 ; Search time 2691.71 Seconds
(without alignments)
10559.003 Million cell updates/sec

Title: US-10-607-806-1_COPY_1_500

Perfect score: 499.6

Sequence: 1 gactactcctgaccttctgt.....accacataagaggactga 500

Scoring table: IDENTITY_NUC

Gapop 10.0 , Gapext 1.0

Searched: 5893141 seqs, 28421725653 residues

Total number of hits satisfying chosen parameters: 11766282

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : GenEmbl.*

1: gb_ba.*
2: gb_in.*
3: gb_env.*
4: gb_om.*
5: gb_ov.*
6: gb_pac.*
7: gb_ph.*
8: gb_pr.*
9: gb_ro.*
10: gb_sts.*
11: gb_sy.*
12: gb_un.*
13: gb_vl.*
14: gb_htg.*
15: gb_pl.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB	ID	Description
C 1	499.6	100.0	122302	8	AC003982	AC003982 Homo sapi
C 2	499.6	100.0	220384	14	AC078926	AC078926 Homo sapi
C 3	218.6	43.8	150453	8	AC015920	AC015920 Homo sapi
4	218.6	43.8	160254	14	AC007763	AC007763 Homo sapi
5	218.6	43.8	163229	8	AC115085	AC115085 Homo sapi
C 6	218.6	43.8	178097	14	AC016704	AC016704 Homo sapi
C 7	217.2	43.5	79223	8	AL391114	AL391114 Human DNA
C 8	216.4	43.3	169566	8	AC026470	AC026470 Homo sapi
9	214.8	43.0	662	6	AX386112	AX386112 Sequence
10	214.8	43.0	123576	8	AC015969	AC015969 Homo sapi
11	214.8	43.0	190306	8	AC146080	AC146080 Pan trogl
C 12	214.6	43.0	90000	8	AF002789	AF002789 Homo sapi
C 13	214.4	42.9	726	10	BV632439	BV632439 S215P6037
C 14	214.4	42.9	193369	14	AC114889	AC114889 Pan trogl
15	214.4	42.9	195782	8	AC074389	AC074389 Homo sapi
16	214.2	42.9	61741	14	AC147308	Continuation (4 of
17	214.2	42.9	110000	14	AC138525	AC138525 Homo sapi
18	214.2	42.9	110000	14	AC147308_2	Continuation (3 of

C 19	214.2	42.9	135758	8	AL136324	AL136324 Human DNA
20	214.2	42.9	191540	8	AC144780	AC144780 Pan trogl
21	214.2	42.9	203842	8	AC010789	AC010789 Homo sapi
22	214	42.8	80068	8	AC008598	AC008598 Homo sapi
C 23	214	42.8	121265	14	AF286885	AF286885 Homo sapi
24	214	42.8	135213	8	AC111149	AC111149 Homo sapi
25	214	42.8	149387	8	AC016266	AC016266 Homo sapi
C 26	214	42.8	174222	14	AP001197	AP001197 Homo sapi
C 27	214	42.8	175553	14	AP001096	AP001096 Homo sapi
C 28	214	42.8	178042	14	AC018745	AC018745 Homo sapi
29	214	42.8	182884	8	AC007996	AC007996 Homo sapi
30	214	42.8	189975	8	AC109512	AC109512 Homo sapi
31	214	42.8	191318	14	AC020670	AC020670 Homo sapi
32	213.8	42.8	4278	8	HSM804155	AL832844 Homo sapi
33	213.8	42.8	136713	8	AC112255	AC112255 Homo sapi
C 34	213.8	42.8	138589	8	AC092966	AC092966 Homo sapi
C 35	213.8	42.8	139953	8	HSBA504H3	AL121585 Human DNA
C 36	213.8	42.8	186649	8	AL929410	AL929410 Human chr
C 37	213.8	42.8	209157	8	CNS01DW4	AL136332 Human chr
38	213.4	42.7	587	6	AX389850	AX389850 Sequence
39	213.4	42.7	59107	14	HS409J21	283824 Homo sapien
C 40	213.4	42.7	89086	8	HS1188J21	AL109963 Human DNA
C 41	213.4	42.7	129266	8	AL731567	AL731567 Human DNA
42	213.4	42.7	153094	14	AC096715	AC096715 Homo sapi
43	213.4	42.7	159123	8	AC007739	AC007739 Homo sapi
44	213.4	42.7	168174	6	AR584462	AR584462 Sequence
45	213.4	42.7	168273	6	AR584403	AR584403 Sequence

ALIGNMENTS

RESULT 1	AC003982	122302 bp	DNA	linear	PRI 18-MAR-1999
LOCUS	Homo sapiens PAC clone 166H1 from 12q, complete sequence.				
DEFINITION	AC003982				
ACCESSION	AC003982.1	GI:2769695			
VERSION	HTG.				
KEYWORDS	Homo sapiens (human)				
SOURCE	Homo sapiens				
ORGANISM	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini; Hominidae; Homo.				
REFERENCE	1 (bases 1 to 122302)				
AUTHORS	Sulston, J.E. and Waterston, R.				
TITLE	Toward a complete human genome sequence				
JOURNAL	Genome Res. 8 (11), 1097-1108 (1998)				
PUBMED	9847074				
REFERENCE	2 (bases 1 to 122302)				
AUTHORS	Bradshaw, H., Wu, X. and Ozersky, P.				
TITLE	The sequence of Homo sapiens PAC clone 166H1				
JOURNAL	Unpublished (1999)				
REFERENCE	3 (bases 1 to 122302)				
AUTHORS	Waterston, R.				
TITLE	Direct Submission				
JOURNAL	Submitted (13-JAN-1998) Department of Genetics, Washington University, 444 Forest Park Avenue, St. Louis, Missouri 63108, USA				
REFERENCE	4 (bases 1 to 122302)				
AUTHORS	Waterston, R.				
TITLE	Direct Submission				
JOURNAL	Submitted (18-MAR-1999) Department of Genetics, Washington University, 444 Forest Park Avenue, St. Louis, Missouri 63108, USA				
COMMENT	SUBMITTED BY: WUGSC Genome Sequencing Center Department of Genetics Washington University St. Louis MO 63108, USA http://genome.wustl.edu/gsc mailto:sapienswatson.wustl.edu				

NOTICE: This sequence may not represent the entire insert of this clone. It may be shorter because we only sequence overlapping


```

/rpt_family="I2"
12999. .13303
/rpt_family="Alu"
Complement (13674. .13752)
/Note="match to EST N80020 (NID: g1242721) za91a08.s1"
Complement (13840. .14195)
/Note="match to EST N80020 (NID: g1242721) za91a08.s1"
14196. .14229
/rpt_family="AT_rich"
14240. .14339
/rpt_family="U6"
14349. .14646
/rpt_family="Alu"
14652. .14940
/rpt_family="Alu"

Query Match
Best Local Similarity 99.8%; Pred. No. 1.9e-142; Length 122302;
Matches 499; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 1 GACCTACCTCGACCTTTGTGCCAGGTTCTTAGCATATGGGACCTGGGATGGAGTTAGCGC 60
Db 33287 GACCTACCTCGACCTTTGTGCCAGGTTCTTAGCATATGGGACCTGGGATGGAGTTAGCGC 33228

QY 61 TCAGTTAATAGTAACCTATTAGCCAGGTCGGTGCTCATCTGTATTCACGACATTT 120
Db 33227 TCAGTTAATAGTAACCTATTAGCCAGGTCGGTGCTCATCTGTATTCACGACATTT 33168

QY 121 GGGAGACCGAGTGGTGGATCACTTGAGACGAGGTTTCAGACCCAGCTGGCCAAACAT 180
Db 33167 GGGAGACCGAGTGGTGGATCACTTGAGACGAGGTTTCAGACCCAGCTGGCCAAACAT 33108

QY 181 GCGAAAAACACTATCTTAATAAAAAATACAAAAATAGCCAGGTCGGTGCTTGCTTA 240
Db 33107 GCGAAAAACACTATCTTAATAAAAAATACAAAAATAGCCAGGTCGGTGCTTGCTTA 33048

QY 241 TAGTCCGAGTACACAGAGGCTGGGGGAGAAGATCACTTGAACCTGGGAGTGGAGGT 300
Db 33047 TAGTCCGAGTACACAGAGGCTGGGGGAGAAGATCACTTGAACCTGGGAGTGGAGGT 32988

QY 301 TGCAGTGAGCCAGATTGCACCTGCATCTGCACCTGCACTGGAAAAAGGTAATTAATAC 360
Db 32987 TGCAGTGAGCCAGATTGCACCTGCATCTGCACCTGCACTGGAAAAAGGTAATTAATAC 32928

QY 361 TTTACTTGCACCACTAGCTGCTTCTCTTCTTTCAGGACCCCAATCACCCTTAGCA 420
Db 32927 TTTACTTGCACCACTAGCTGCTTCTCTTCTTTCAGGACCCCAATCACCCTTAGCA 32868

QY 421 TCCTTCAGGCTAAAYCTAGGAGCAGTGCCTGCTCTGCTTGTATGACCCCAAGGA 480
Db 32867 TCCTTCAGGCTAAAYCTAGGAGCAGTGCCTGCTCTGCTTGTATGACCCCAAGGA 32808

QY 481 ACCACATAGAGGACTGA 500
Db 32807 ACCACATAGAGGACTGA 32788

RESULT 2
AC078926/c
LOCUS AC078926 220384 bp DNA linear HTG 26-MAR-2002
DEFINITION Homo sapiens chromosome 12 clone RP11-836M11, WORKING DRAFT
SEQUENCE, 11 unordered pieces.
ACCESSION AC078926
VERSION AC078926.23 GI:19718703
KEYWORDS HTG; HTGS_PHASE1; HTGS_DRAFT; HTGS_FULLTOP.
SOURCE Homo sapiens
ORGANISM Homo sapiens
Eukaryota; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini;
Hominoidea; Homo.
REFERENCE 1 (bases 1 to 220384)
Muzny,D.M., Adams,C., Adio-Oduola,B., Ali-osman,F.R., Allen,C.,
Alsbrooks,S.L., Amaratunge,H.C., Are,J.R., Ayele,M., Banks,T.,

```

Barbaria, J., Benton, J., Bimage, K., Blankenburg, K., Bonnin, D., Bouck, J., Bowie, S., Brieva, M., Brown, E., Brown, M., Bryant, N.P., Buhay, C., Burch, P., Burkett, C., Burrell, K.L., Byrd, N.C., Caron, T., Carter, M., Cavazos, S.R., Chacko, J., Chavez, D., Chen, G., Chen, R., Chen, Z., Chowdhry, I., Christopoulos, C., Cleveland, C.D., Cox, C., Coyle, M.D., Dathorne, S.R., David, R., Davila, M.L., Davis, C., Davy-Carroll, L., Dederich, D.A., Delaney, K.R., Delgado, O., Denn, A.L., Ding, Y., Dinh, H.H., Douthwaite, K.J., Draper, H., Dugan-Rocha, S., Durbin, K.J., Earnhart, C., Edgar, D., Edwards, C.C., Elhaj, C., Escotto, M., Falls, T., Ferraguto, D., Flagg, N., Ford, J., Foster, P., Frantz, P., Gabisi, A., Gao, J., Garcia, A., Garner, T., Garza, N., Gill, R., Gorrell, J.H., Guevara, W., Gunaratne, P., Hale, S., Hamilton, K., Harris, C., Harris, K., Hart, M., Havlak, P., Hayes, A., Hernandez, J., Hernandez, O., Hodgson, A., Hoques, M., Holloway, C., Hollins, B., Honsi, F., Howard, S., Huber, J., Hulyk, S., Hume, J., Jackson, L.E., Jacobson, B., Jia, Y., Johnson, R., Jolivet, S., Joudah, S., Karlsson, E., Kelly, S., Khan, U., King, L., Korvah, J., Kovar, C., Kratovic, J., Kureshi, A., Landry, N., Leal, B., Lewis, L.C., Lewis, L., Li, J., Li, Z., Lichtarge, O., Liu, C., Liu, J., Liu, W., Loulseg, H., Lozano, R.J., Lu, X., Lucier, A., Lucier, R., Luna, R., Ma, J., Maheshwari, M., Mapua, P., Martin, R., Martindale, A., Martinez, E., Massey, E., Mawney, E., McLeod, M.P., Meador, M., Mei, G., Metzker, M., Miner, G., Miner, Z., Mitchell, T., Mohabbat, K., Morgan, M., Morris, S., Moser, M., Neal, D., Newton, J., Newton, N., Nguyen, A., Nguyen, N., Nguyen, N., Nickerson, E., Nwokenko, S., Ogih, M., Okwuonu, G., Ogunye, N., Oviedo, R., Pace, A., Payton, B., Peery, J., Perez, L., Peters, L., Pickens, R., Primus, E., Pu, L.L., Quiles, M., Ren, X., Rives, M., Rojas, A., Rojubokan, I., Rolfe, M., Ruiz, S., Savery, G., Scherer, S., Scott, G., Shen, H., Shoshitari, N., Sisson, I., Sodergren, E., Sonaika, T., Sparks, A., Stanley, H., Stone, H., Sutton, A., Svatek, A., Tabor, P., Tamerisa, A., Tamerisa, K., Tang, H., Tansey, J., Taylor, C., Taylor, T., Telford, B., Thomas, N., Thomas, S., Usmani, K., Vasquez, L., Vera, V., Villalon, D., Vinson, R., Wang, Q., Wang, S., Ward-Moore, S., Warren, R., Washington, C., Watlington, S., Williams, G., Williamson, A., Wleczyk, R., Wooden, S., Worley, K., Wu, C., Wu, Y., Zhou, J., Zorrilla, S., Nelson, D., Weinstock, G., and Gibbs, R.

Direct Submission
Unpublished
2 (bases 1 to 220384)
Worley, K.C.
Direct Submission
Submitted (11-AUG-2000) Human Genome Sequencing Center, Department of Molecular and Human Genetics, Baylor College of Medicine, One Baylor Plaza, Houston, TX 77030, USA
3 (bases 1 to 220384)
Worley, K.C.
Direct Submission
Submitted (26-MAR-2002) Human Genome Sequencing Center, Department of Molecular and Human Genetics, Baylor College of Medicine, One Baylor Plaza, Houston, TX 77030, USA
On Mar 26, 2002 this sequence version replaced gi:18449664.

----- Genome Center
Center: Baylor College of Medicine
Center code: BCM
Web site: <http://www.hgsc.bcm.tmc.edu/>
Contact: hgsc-help@bcm.tmc.edu
----- Project Information
Center project name: HCAP
Center clone name: RP11-836M11
----- Summary Statistics
Sequencing vector: Plasmid;
Sequencing vector: M13;
Chemistry: Dye-primer Bodipy: 47% of reads
Chemistry: Dye-terminator Big Dye: 53% of reads
Assembly program: Phrap; version 0.990329
Consensus quality: 234783 bases at least Q40
Consensus quality: 241821 bases at least Q30
Consensus quality: 245671 bases at least Q20
Estimated insert size: 21987; sum-of-contigs estimation
Quality coverage: 8.9x in Q20 bases; sum-of-contigs estimation

repeat_region /rpt_family="AluSq/x" 29748. .30037
repeat_region /rpt_family="AluSx" complement(30051. .30397)
repeat_region /rpt_family="MLT1A2" 30398. .30446
repeat_region /rpt_family="L1MB4A" 30753. .31064
repeat_region /rpt_family="AluSx" 31124. .31231
repeat_region /rpt_family="L1MB4A"

Query Match 43.8%; Score 218.6; DB 8; Length 150453;
Best Local Similarity 83.5%; Pred. No. 5.4e-56;
Matches 248; Conservative 0; Mismatches 49; Indels 0; Gaps 0;

QY 65 TTAATAGTAATCATACCAAGTGGCGTCTATCTCTATATCCAGCACATTTGGGA 124
Db 40951 TTAATAGTAATCATATGCGCGGTGTGTGCTCACACCTGTATCCAGCACATTTGGGA 40892
QY 125 GACCCAGTTGGTGATCACTTGAGCAGGAGTTTGAGACAGCAGCTGCCAACATGGCA 184
Db 40891 GCCCAGAGGTTGATCACTGAGTGAGGTTTGAGACAGCAGCTGCCAACATGGCA 40832
QY 185 AAACACTATCTATAAATAATACAAAATTAGCCAGTGTTGGTGACCTTGCCTATAGT 244
Db 40831 AAACCCATCTCTACTATAAATAATACAAAATTAGCCAGGCGTGTGTGTCATGCTGTAA 40772
QY 245 CCACGTACACAGGAGGTTGGCGGAGAGATCACTTGAACCTGGAGGTGGAGTTTCCA 304
Db 40771 CCACGTACTCTGGAGGCTGGAGGAGGAGATCACTTGAACCTGGAGGCTGGAGTTTCCA 40712
QY 305 GTGAGCCAGATTGACCACTCACTCCAGCCTCGGAAAAAGGGTAATTAATACT 361
Db 40711 GTGAGCCGAGATCACACCACTGCTCCAGCCTGGGCAACGAGAGTCTGTCT 40655

AC007763 160254 bp DNA linear HTG 18-NOV-2001
Homo sapiens clone RP11-110D17, WORKING DRAFT SEQUENCE, 11
unordered pieces.
AC007763
AC007763.6 GI:16974272
HTG: HTGS_PHASE1; HTGS_DRAFT; HTGS_FULLTOP.
Homo sapiens (human)
Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini;
Homnidae; Homo.
1 (bases 1 to 160254)
Birren,B., Linton,L., Nusbaum,C. and Lander,E.
Homo sapiens chromosome, clone RP11-110D17
Unpublished
2 (bases 1 to 160254)
Birren,B., Linton,L., Nusbaum,C., Lander,E., Allen,N., Anderson,M.,
Baker,J., Baldwin,J., Barna,N., Beckerly,R., Benn,J., Brown,A.,
Castle,A., Cerny,J., Colangelo,M., Collins,S., Collymore,A.,
Cooke,P., DeArillano,K., Depayre,E., Devon,K., Dewar,K.,
Donelan,L., Doyle,M., Ferreira,P., FitzHugh,W., Forrest,C.,
Funke,R., Gage,D., Galagan,J., Gardyna,S., Gilbert,D., Grant,G.,
Hagos,B., Heaford,A., Horton,L., Howland,J.C., Jones,C., Kann,L.,
Karatas,A., Lebecky,J., Lie,C., Locke,K., Macdonald,P.,
Marquis,N., McEwan,P., McGurk,A., McKernan,K., McLaughlin,J.,
Meldrim,J., Molla,M., Morris,W., Morrow,J., Mychaleckyj,J.,
Naylor,J., Niloff,M., O'Connor,F., O'Donnell,P., Pavlin,B.,
Peterson,K., Pollara,V., Riley,R., Roberts,D., Roy,A., Severy,P.,
Stange-Thomann,N., Stojanovic,N., Stone,C., Subramanian,A.,
Tessaye,S., Torruella-Miller,I., Vassiliev,H., Vo,A., Wagner,A.,
Wheeler,J., Wu,X., Wyman,D., Ye,W.J. and Zody,M.
Direct Submission
Submitted (08-JUN-1999) Whitehead Institute/MIT Center for Genome
Research, 320 Charles Street, Cambridge, MA 02141, USA

On Nov 18, 2001 this sequence version replaced gi:14030080.
All repeats were identified using RepeatMasker:
Smit, A.P.A. & Green, P. (1996-1997)
http://ftp.genome.washington.edu/RM/RepeatMasker.html
----- Genome Center
Center: Whitehead Institute/ MIT Center for Genome Research
Center code: WIBR
Web site: http://www-seq.wi.mit.edu
Contact: sequence_submissions@genome.wi.mit.edu
----- Project Information
Center project name: L854
Center clone name: 110 D 17
----- Summary Statistics
Sequencing vector: M13; M77815; 21% of reads
Chemistry: Dye-primer-amersham; n/a; 79% of reads
Chemistry: Dye-terminator Big Dye; 82% of reads
Assembly program: Phrap; version 0.960731
Consensus quality: 156131 bases at least Q40
Consensus quality: 157776 bases at least Q30
Consensus quality: 158528 bases at least Q20
Insert size: 90000; agarose-fp
Quality coverage: 17.8 in Q20 bases.
NOTE: This is a 'working draft' sequence. It currently
* consists of 11 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.
* 1 20714: contig of 20714 bp in length
* 20815 20814: gap of 100 bp
* 20815 21860: contig of 1046 bp in length
* 21861 21960: gap of 100 bp
* 21961 23124: contig of 1164 bp in length
* 23125 23224: gap of 100 bp
* 23225 25014: contig of 1790 bp in length
* 25015 25114: gap of 100 bp
* 25115 26473: contig of 1359 bp in length
* 26474 26573: gap of 100 bp
* 26574 31375: contig of 4802 bp in length
* 31376 31475: gap of 100 bp
* 31476 40242: contig of 8767 bp in length
* 40243 40342: gap of 100 bp
* 40343 55414: contig of 15072 bp in length
* 55415 55514: gap of 100 bp
* 55515 76039: contig of 20525 bp in length
* 76040 76139: gap of 100 bp
* 76140 110758: contig of 34619 bp in length
* 110759 110858: gap of 100 bp
* 110859 160254: contig of 49396 bp in length.
Location/Qualifiers
1. 160254
/organism="Homo sapiens"
/mol_type="genomic DNA"
/db_xref="taxon:9606"
/clone="RP11-110D17"
/clone_lib="RPC1-11 Human Male BAC"
1. 20714
/note="assembly_fragment"
clone_end:SP6
vector_side:left
20715. 20814
/estimated_length=100
20815. 21860
/note="assembly_fragment"
21861. 21960
/estimated_length=100
21961. 23124
/note="assembly_fragment"
23125. 23224

COMMENT

FEATURES
source

misc_feature

gap

misc_feature

gap

misc_feature

gap

```

/misc_feature      /estimated_length=100
23225. .25014
/note="assembly_fragment"
gap               /estimated_length=100
25015. .25114
/misc_feature      /estimated_length=100
25115. .26473
/note="assembly_fragment"
gap               /estimated_length=100
26474. .26573
/misc_feature      /estimated_length=100
26574. .31375
/note="assembly_fragment"
gap               /estimated_length=100
31376. .31475
/misc_feature      /estimated_length=100
31476. .40242
/note="assembly_fragment"
gap               /estimated_length=100
40243. .40342
/misc_feature      /estimated_length=100
40343. .55414
/note="assembly_fragment"
gap               /estimated_length=100
55415. .55514
/misc_feature      /estimated_length=100
55515. .76039
/note="assembly_fragment"
gap               /estimated_length=100
76040. .76139
/misc_feature      /estimated_length=100
76140. .110758
/note="assembly_fragment"
gap               /estimated_length=100
110759. .110858
/misc_feature      /estimated_length=100
110859. .160254
/note="assembly_fragment"

ORIGIN
Query Match      43.8%; Score 218.6; DB 14; Length 160254;
Best Local Similarity 83.5%; Pred No. 5.3e-56;
Matches 248; Conservative 0; Mismatches 49; Indels 0; Gaps 0;

Qy 65 TTAATAGTAACCTATTAGCCAGGTGCGGTGCTCATGTCTGTATTCACGACCTTTGGGA 124
Db 46713 TTAATAGTAATATCATGCGCGGTGTGTGCTCACACCTGTATTCACGACCTTTGGGA 46772

Qy 125 GACCGAGTGGTGGATCATTGAGCAGCAGAGTTTGAGCAGCCTGGCCAAATGGCA 184
Db 46773 GGCCGAGAAGGTTGGATCACCTGAGTGCAGGAGTTGAGCAGCCTGGCCAAATGGCA 46832

Qy 185 AAACATATCTCTAATAAATAACAATAATAGCAGGTGTGGTGGCCTGCTATAGT 244
Db 46833 AAACCCATCTCTAATAAATAACAATAATAGCAGGTGTGGTGGCCTGCTATAGT 46892

Qy 245 CCCAGCTACACAGGAGTGGGGCAGAGAAATCACTTGAACCTGGGAGGTGGAGTTGCA 304
Db 46893 CCCAGCTACTCTGGAGCTGAGGAGGAGATCACTTGAACCTGGGAGGTGGAGTTGCA 46952

Qy 305 GTGAGCCGAAGTTGCACACTGCACTCCAGCTCGAATAAAGGTTAATTAATCACT 361
Db 46953 GTGAGCCGAGATCACACACTGCACTCCAGCTCGGCAACAAGAGATTCCTGCT 47009

RESULT 5
AC115085 Homo sapiens 163229 bp DNA linear PRI 01-AUG-2002
LOCUS Homo sapiens chromosome 17, clone RP11-280F10, complete sequence.
DEFINITION AC115085
ACCESSION AC115085
VERSION AC115085.4 GI:21314912
KEYWORDS HTG.
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
Eukaryota; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini;
Hominae; Homo.
REFERENCE 1 (bases 1 to 163229)
AUTHORS Birren,B., Nusbaum,C. and Lander,E.
TITLE Homo sapiens chromosome 17, clone RP11-280F10

```

JOURNAL REFERENCE AUTHORS

Unpublished
2 (bases 1 to 163229)
Birren,B., Linton,L., Nusbaum,C., Lander,E., Ali,A., Allen,N.,
Anderson,S., Barna,N., Bastien,V., Bloom,T., Boguslavsky,L.,
Boukhalter,B., Brown,A., Camarata,J., Campopiano,A., Chang,J.,
Chazaro,B., Choepel,Y., Colangelo,M., Collins,S., Collymore,A.,
Cook,A., Cooke,P., DeArellano,K., Dewar,K., Diaz,J.S., Dodge,S.,
Faroo,S., Ferreira,P., FitzHugh,W., Gage,D., Galagan,J., Gardyna,S.,
Ginde,S., Gord,S., Goyette,M., Graham,L., Grand-Pierre,N.,
Hagos,B., Horton,L., Hulme,W., Iliev,I., Johnson,R., Jones,C.,
Kamat,A., Karatas,A., Kells,C., Lacombe,K., Lamazares,R.,
Landers,T., Lehoczy,J., Levine,R., Lindblad-Toh,K., Liu,G.,
MacLean,C., Macdonald,P., Major,J., Marquis,N., Matthews,C.,
McCarthy,M., McEwan,P., McKernan,K., Meldrum,J., Meneus,L.,
Mihova,T., Mlengwa,V., Murphy,T., Naylor,J., Nguyen,C., Nicol,R.,
Norbu,C., Norman,C.H., O'Connor,T., O'Donnell,P., O'Neill,D.,
Oliver,J., Peterson,K., Phunkhang,P., Pierre,N., Pollara,V.,
Raymond,C., Retta,R., Rieback,M., Riley,R., Rise,C., Rogov,P.,
Roman,J., Rosetti,M., Roy,A., Santos,R., Schauer,S., Schupback,R.,
Seaman,S., Severy,P., Spencer,B., Stange-Thomann,N., Stojanovic,N.,
Strauss,N., Subramanian,A., Talamas,J., Tesfaye,S., Theodore,J.,
Topham,K., Travers,M., Travis,N., Trigilio,J., Vassiliev,H.,
Viel,R., Vo,A., Wilson,B., Wu,X., Wyman,D., Ye,W.J., Young,G.,
Zainoun,J., Zembek,L., Zimmer,A. and Zody,M.

TITLE JOURNAL

Submitted (14-WAR-2002) Whitehead Institute/MIT Center for Genome
Research, 320 Charles Street, Cambridge, MA 02141, USA
3 (bases 1 to 163229)

REFERENCE AUTHORS

Birren,B., Linton,L., Nusbaum,C., Lander,E., Ali,A., Allen,N.,
Anderson,S., Barna,N., Bastien,V., Bloom,T., Boguslavsky,L.,
Boukhalter,B., Brown,A., Camarata,J., Campopiano,A., Chang,J.,
Chazaro,B., Choepel,Y., Colangelo,M., Collins,S., Collymore,A.,
Cook,A., Cooke,P., DeArellano,K., Dewar,K., Diaz,J.S., Dodge,S.,
Faroo,S., Ferreira,P., FitzGerald,M., FitzHugh,W., Gage,D.,
Galagan,J., Gardyna,S., Ginde,S., Gord,S., Goyette,M., Graham,L.,
Grand-Pierre,N., Hagos,B., Horton,L., Hulme,W., Iliev,I.,
Johnson,R., Jones,C., Kamat,A., Karatas,A., Kells,C., LaRocque,K.,
Lamazares,R., Landers,T., Lehoczy,J., Levine,R., Lindblad-Toh,K.,
Liu,G., MacLean,C., Macdonald,P., Major,J., Marquis,N.,
Matthews,C., McCarthy,M., McEwan,P., McKernan,K., Meldrum,J.,
Meneus,L., Mihova,T., Mlengwa,V., Murphy,T., Naylor,J., Nguyen,C.,
Nicol,R., Norbu,C., Norman,C.H., O'Connor,T., O'Donnell,P.,
O'Neill,D., Oliver,J., Peterson,K., Phunkhang,P., Pierre,N.,
Pollara,V., Raymond,C., Retta,R., Rieback,M., Riley,R., Rise,C.,
Rogov,P., Roman,J., Rosetti,M., Roy,A., Santos,R., Schauer,S.,
Schupback,R., Seaman,S., Severy,P., Spencer,B., Stange-Thomann,N.,
Stojanovic,N., Strauss,N., Subramanian,A., Talamas,J., Tesfaye,S.,
Theodore,J., Topham,K., Travers,M., Travis,N., Trigilio,J.,
Vassiliev,H., Viel,R., Vo,A., Wilson,B., Wu,X., Wyman,D., Ye,W.J.,
Young,G., Zainoun,J., Zembek,L., Zimmer,A. and Zody,M.

TITLE JOURNAL

Submitted (03-JUN-2002) Whitehead Institute/MIT Center for Genome
Research, 320 Charles Street, Cambridge, MA 02141, USA
4 (bases 1 to 163229)

REFERENCE AUTHORS

Birren,B., Nusbaum,C., Lander,E., Ali,A., Allen,N., Anderson,S.,
Barna,N., Bastien,V., Bloom,T., Boguslavsky,L., Boukhalter,B.,
Camarata,J., Chang,J., Chazaro,B., Choepel,Y., Collymore,A.,
Cook,A., Cooke,P., DeArellano,K., Dewar,K., Diaz,J.S., Dodge,S.,
Faroo,S., Ferreira,P., FitzGerald,M., Gage,D., Galagan,J.,
Gardyna,S., Gord,S., Graham,L., Grand-Pierre,N., Hagos,B.,
Horton,L., Hulme,W., Iliev,I., Johnson,R., Jones,C., Kamat,A.,
Karatas,A., Kells,C., Landers,T., Levine,R., Lindblad-Toh,K.,
Liu,G., MacLean,C., Macdonald,P., Major,J., Matthews,C.,
McCarthy,M., Meldrum,J., Meneus,L., Mihova,T., Mlengwa,V.,
Murphy,T., Naylor,J., Nguyen,C., Nicol,R., Norbu,C., Norman,C.H.,
O'Connor,T., O'Donnell,P., O'Neill,D., Oliver,J., Peterson,K.,
Phunkhang,P., Pierre,N., Raymond,C., Retta,R., Rise,C., Rogov,P.,
Roman,J., Roy,A., Schauer,S., Schupback,R., Seaman,S., Severy,P.,
Smith,C., Spencer,B., Stange-Thomann,N., Stojanovic,N., Talamas,J.,
Tesfaye,S., Theodore,J., Topham,K., Travers,M., Vassiliev,H.,
Viel,R., Vo,A., Wilson,B., Wu,X., Wyman,D., Young,G., Zainoun,J.,
Zembek,L., Zimmer,A. and Zody,M.

TITLE

Direct Submission

JOURNAL
COMMENT

Submitted (01-AUG-2002) Whitehead Institute/MIT Center for Genome Research, 320 Charles Street, Cambridge, MA 02141, USA
On Jun 3, 2002 this sequence version replaced gi:21218421.
All repeats were identified using RepeatMasker:
Smit, A.F.A. & Green, P. (1996-1997)
<http://ftp.genome.washington.edu/RM/RepeatMasker.html>
----- Genome Center
Center: Whitehead Institute/ MIT Center for Genome Research
Center code: WIBR
Web site: <http://www-seq.wi.mit.edu>
Contact: sequence_submissions@genome.wi.mit.edu
----- Project Information
Center project name: L26095
Center clone name: 280 F 10

FEATURES	Source	Location/Qualifiers
repeat_region	1. .163229	/clone_lib="RPC1-11 Human Male BAC"
repeat_region	18. .71	/rpt_family="MER58A"
repeat_region	complement(108. .412)	/rpt_family="AluSx"
repeat_region	413. .579	/rpt_family="MER58A"
repeat_region	676. .811	/rpt_family="FLAM_C"
repeat_region	814. .844	/rpt_family="AT_rich"
repeat_region	complement(2166. .2488)	/rpt_family="AluUb"
repeat_region	3244. .3536	/rpt_family="AluSg"
repeat_region	complement(3628. .3818)	/rpt_family="Tigger1"
repeat_region	complement(3819. .4141)	/rpt_family="AluSx"
repeat_region	complement(4143. .4303)	/rpt_family="PRAM"
repeat_region	complement(4304. .5368)	/rpt_family="Tigger1"
repeat_region	5369. .5400	/rpt_family="(TC) n"
repeat_region	complement(5401. .5418)	/rpt_family="Tigger1"
repeat_region	complement(5419. .5665)	/rpt_family="AluUb"
repeat_region	complement(5666. .6204)	/rpt_family="Tigger1"
repeat_region	complement(6205. .6507)	/rpt_family="AluSx"
repeat_region	complement(6508. .6701)	/rpt_family="Tigger1"
repeat_region	complement(6702. .7009)	/rpt_family="AluSg"
repeat_region	complement(7010. .7497)	/rpt_family="Tigger1"
repeat_region	7756. .8073	/rpt_family="AluSg"
repeat_region	8780. .8968	/rpt_family="L1MA6"
repeat_region	9007. .9303	/rpt_family="AluSx"
repeat_region	10034. .10259	/rpt_family="WIP"
repeat_region	complement(10428. .10732)	/rpt_family="AluSx"
repeat_region	11106. .11392	

repeat_region	/rpt_family="AluSx"
repeat_region	11402. .11506
repeat_region	/rpt_family="(TA)n"
repeat_region	11510. .11554
repeat_region	/rpt_family="LTR66"
repeat_region	11573. .11611
repeat_region	/rpt_family="(TAGA)n"
repeat_region	complement(11727. .12008)
repeat_region	/rpt_family="AluSq"
repeat_region	complement(12137. .12267)
repeat_region	/rpt_family="FLAM_A"
repeat_region	12355. .12376
repeat_region	/rpt_family="AT_rich"
repeat_region	complement(12936. .13228)
repeat_region	/rpt_family="AluSq"
repeat_region	13378. .13691
repeat_region	/rpt_family="11PB1"
repeat_region	15066. .15372
repeat_region	/rpt_family="AluSq"
repeat_region	15539. .15843
repeat_region	/rpt_family="AluSx"
repeat_region	16005. .16067
repeat_region	/rpt_family="(TTCA)n"
repeat_region	16113. .16413
repeat_region	/rpt_family="AluSx"
repeat_region	16692. .17434
repeat_region	/rpt_family="L1MB2"
repeat_region	complement(17491. .17782)
repeat_region	/rpt_family="AluSc"
repeat_region	complement(18043. .18201)
repeat_region	/rpt_family="MER20"
repeat_region	complement(18219. .18321)
repeat_region	/rpt_family="MER94"
repeat_region	18327. .18445
repeat_region	/rpt_family="L2"
repeat_region	complement(18446. .18757)
repeat_region	/rpt_family="AluY"
repeat_region	18758. .19077
repeat_region	/rpt_family="L2"
repeat_region	complement(19078. .19376)
repeat_region	/rpt_family="AluSq"

Query Match	43.8%;	Score 218.6;	DB 8;	Length 163229;
Best Local Similarity	83.5%;	Pred. No. 5.3e-56;		
Matches 248;	Conservative	0;	Mismatches 49;	Indels 0; Gaps 0;
Qy 65	TTAATAAGTAACTCATTAGCCAGGTGCGGTGCTCATGTCTGTATATCCAGCACATTTGGGA	124		
Db 50158	TTAATAAGTAAATCATGCGCGGTGTGTGGCTCACACCTGTAAATCCAGCACATTTGGGA	50217		
Qy 125	GACCGAGTTGGTGGATCATTTGAGAGCAGAGTTTGAGACCAGCCTGGCCCAACATGGCA	184		
Db 50218	GGCCGAGAAGGCTGGATCACTTGAGCTCAGGAGTTTGAGACCGCCTGGCCCAACATGGCA	50277		
Qy 185	AAACACTATCTCTAATAAATAACAAAAATATAGCCAGGTGTGTGGGCACTTGCCTATAGT	244		
Db 50278	AAACCCCATCTCTACTAAATAATACAAAAATATAGCCAGCGTGTGTGTGCTGCTGTAAT	50337		
Qy 245	CCCAGCTACACAGGAGGCTGGGGCAGAGAATACTTGAACCTGGAGGTGGAGGTTGCA	304		
Db 50338	CCCAGCTACTCTGGAGGCTTGAGCCAGGAGAATCACTTGAACCTGGAGGCTGAGGTTCCA	50397		
Qy 305	GTGAGCCAGATTGACCACTGCATCCAGCCTGCAAAAAAGGTAATTAATAACT	361		
Db 50398	GTGAGCCGAGATCACACCCTGCATCCAGCCTGGGCAACAAAGCGAGTTCTGTCT	50454		

RESULT 6

AC016704/c

LOCUS AC016704 178097 bp DNA linear HTG 07-JUL-2000

DEFINITION Homo sapiens chromosome 17 clone RP11-202G12, WORKING DRAFT

SEQUENCE, 20 unordered pieces.

ACCESSION AC016704

RESULT 6					
AC016704/c					
LOCUS	AC016704	178097 bp	DNA	linear	HTG 07-JUL-2000
DEFINITION	Homo sapiens chromosome 17 clone RP11-202G12, WORKING DRAFT				
	SEQUENCE, 20 unordered pieces.				
ACCESSION	AC016704				

complement (55420. .55580), complement (54120. .54169),
complement (50780. .50966), complement (49799. .49965),
complement (44842. .44954), complement (44249. .44352),
complement (35749. .35886), complement (32782. .32946),
complement (28849. .28981), complement (23938. .24204),
complement (21361. .21572), complement (11346. .11522),
complement (9355. .9426), complement (6975. .7065),
complement (AL365505.15:97507. .97655),
complement (AL365505.15:93798. .93962),
complement (AL365505.15:86445. .89025))
/gene="RBL1"
/locus_tag="RP11-382A12.1-001"
/product="retinoblastoma-like 1 (p107)"
/note="match: ESTs: A1128307.1 BB624672.1 BB870263.1
BE332924.1 BE650226.1 BF398744.1 BG387892.1 BM463052.1
BQ004524.1 BU058420.1 BU178775.1 BU556640.1 T85305.1
match: cDNAs: Em: L14812.1 Em: M74547.1"
Join (complement (AL36172.16:5346. .5562),
complement (77685. .77818), complement (56682. .56882),
complement (55929. .55993), complement (55689. .55817),
complement (55420. .55580), complement (54120. .54169),
complement (50780. .50966), complement (49799. .49965),
complement (44842. .44954), complement (44249. .44352),
complement (35749. .35886), complement (32782. .32946),
complement (28849. .28981), complement (23938. .24204),
complement (21361. .21572), complement (11346. .11522),
complement (9355. .9426), complement (6975. .7065),
complement (AL365505.15:97507. .97655),
complement (AL365505.15:93670. .93962))
/gene="RBL1"
/locus_tag="RP11-382A12.1-001"
/product="retinoblastoma-like 1 (p107)"
/note="match: ESTs: AA312113.1 AL732304.1 AW209660.1
AW354050.1 AW951018.1 BB650226.1 BE854544.1 BG387892.1
BM463052.1 BQ999505.1 BU178775.1
match: cDNAs: BC032247.1"
Join (complement (AL36172.16:5346. .5501),
complement (77685. .77818), complement (56682. .56882),
complement (55929. .55993), complement (55689. .55817),
complement (55420. .55580), complement (54120. .54169),
complement (50780. .50966), complement (49799. .49965),
complement (44842. .44954), complement (44249. .44352),
complement (35749. .35886), complement (32782. .32946),
complement (28849. .28981), complement (23938. .24204),
complement (21361. .21572), complement (11346. .11522),
complement (9355. .9426), complement (6975. .7065),
complement (AL365505.15:97507. .97655),
complement (AL365505.15:88855. .89025))
/gene="RBL1"
/locus_tag="RP11-382A12.1-001"
/standard_name="OTTHUMP00000030892"
/note="match: proteins: O55081 P28749 Q08999 Q64700
Q64701"
/codon_start=1
/product="retinoblastoma-like 1 (p107)"
/protein_id="CAI95151.1"
/db_xref="GI:66347743"
/translation="MFEDKPHAEGAAVVAAGAEALQALCOELNLDGSAEALDDFTA
IRGNSLEGVTHWACSLYVACRSIIPTVKGIMEGNCVSLTRILRSAKLSIQFF
SKMKWMDNMLPOEFERIERLERNPEVSTVI PKYEPIFLDI FQNPYBEPKLP
RKORRIPCSVKDLNFQWTLFVYTKGNFRMGDDLVNSYHLLCCLDLIPANA
IMCPN RODLLNPSKGLSDFTADFTASEEPPCIIAVLCELHDGLLVEAKG
KEHYFKPYIS KLFDRKILKGECLLDLSFTDNSKAVNKEEYVLTVDGDERIF
FLGADAEIEGTFR KFTDRTPGLKTAQANVEYNLQOHFEKRSFAPSTPLTGR
LYREKEAVITPVASATQ SVSRLOSIVAGLNKAPSDQLINIFESCV
RNPVENIMKLGIEODIFHRSLMACCLEI GSHIDFVNRLLKLAELIYK
ILETVMVOETRRLLHGMDSVLLQODIFHRSLMACCLEI VLFAYS
SPRTFPWIIENVLQPFYKVI EWIRSERGLSRDMVKHLNSEEQIL
LESL AWSHDSALWEALQVSNKVPCEVIFPNPFNGTNGNGVOGHLPL
MPMSPLMHPRVKE VRTDGSRLRDMQPLSPISVHERYSSPTAGS
AKRRLFGEDPPKEMLMDKIIITGCKLK IAPSSSITAEVNSILPQ
GTLTMTATPVTGTHKVTIPLHGVA NDAGEITLPLSNV TNOESK
VSPVSLTAHSLIGASPKQTNLTKAQEVHSTGINRKRKTGSLALFYRKVYH
L

mrna

CDS

EERTFQIMKSYRNQPOANSHYRSVLLKSIPIREVAVYKNKINDDFEMIDCDLEDATK
TPDCSSGPVKEERDGLIKFYNTIYVGRVSKPALKYDLANODHMDAPLSPPPHIKQO
PGSPRRISQOHSIYISPHKNGSLTPRSALYKFGSPSKSLKDNINMIROGEQRTTK
RVIAIDSDAESPARKVQENDDVLLKRLQDVVSEANH"
Join (complement (AL136172.16:5346. .5501),
complement (77685. .77818), complement (56682. .56882),
complement (55929. .55993), complement (55689. .55817),
complement (55420. .55580), complement (54120. .54169),
complement (50780. .50966), complement (49799. .49965),
complement (44842. .44954), complement (44249. .44352),
complement (35749. .35886), complement (32782. .32946),
complement (28849. .28981), complement (23938. .24204),
complement (21361. .21572), complement (11346. .11522),
complement (9355. .9426), complement (6975. .7065),
complement (AL365505.15:97507. .97655),
complement (AL365505.15:93789. .93962))
/gene="RBL1"
/locus_tag="RP11-382A12.1-001"
/standard_name="OTTHUMP00000030893"
/codon_start=1
/product="retinoblastoma-like 1 (p107)"
/protein_id="CAI95152.1"
/db_xref="GI:66347744"
/translation="MFEDKPHAEGAAVVAAGAEALQALCOELNLDGSAEALDDFTA
IRGNSLEGVTHWACSLYVACRSIIPTVKGIMEGNCVSLTRILRSAKLSIQFF
SKMKWMDNMLPOEFERIERLERNPEVSTVI PKYEPIFLDI FQNPYBEPKLP
RKORRIPCSVKDLNFQWTLFVYTKGNFRMGDDLVNSYHLLCCLDLIPANA
IMCPN RODLLNPSKGLSDFTADFTASEEPPCIIAVLCELHDGLLVEAKG
KEHYFKPYIS KLFDRKILKGECLLDLSFTDNSKAVNKEEYVLTVDGDERIF
FLGADAEIEGTFR KFTDRTPGLKTAQANVEYNLQOHFEKRSFAPSTPLTGR
LYREKEAVITPVASATQ SVSRLOSIVAGLNKAPSDQLINIFESCV
RNPVENIMKLGIEODIFHRSLMACCLEI GSHIDFVNRLLKLAELIYK
ILETVMVOETRRLLHGMDSVLLQODIFHRSLMACCLEI VLFAYS
SPRTFPWIIENVLQPFYKVI EWIRSERGLSRDMVKHLNSEEQIL
LESL AWSHDSALWEALQVSNKVPCEVIFPNPFNGTNGNGVOGHLPL
MPMSPLMHPRVKE VRTDGSRLRDMQPLSPISVHERYSSPTAGS
AKRRLFGEDPPKEMLMDKIIITGCKLK IAPSSSITAEVNSILPQ
GTLTMTATPVTGTHKVTIPLHGVA NDAGEITLPLSNV TNOESK
VSPVSLTAHSLIGASPKQTNLTKAQEVHSTGINRKRKTGSLALFYRKVYH
L

	Query Match	43.5%	Score 217.2;	DB 8;	Length 79223;
	Best Local Similarity	85.0%	Pred. No. 1.5e-55;		
	Matches	243;	Conservative	0;	Mismatches 43;
				Indels	0;
				Gaps	0;
Qy	60	CTCAGTTAATAGTAACCTATTAGCCAGGTGGTGGCTCATGTCTGTATTCCAGCACTT	119		
Db	64472	CTCTAGAAAGTGAACAAATTAGCCCGTGCAGTGGCTCACACCTGTATCCAGCACTT	64413		
Qy	120	TGGGAGACCCAGGTTGGGTGGATCATTGAGAGAGAGGATTTGAGACACAGCTTGCCCAACA	179		
Db	64412	TGGGAGGCCGAGGTGGGTGGATCACCCTGAGGTTCAGACCACTTGAGACCACTTGCCCAACA	64353		
Qy	180	TGGCAAAACACTATCTCTAATAAATAACAAAATAGCCAGGTGGTGGCACTTCGCT	239		
Db	64352	TGGCAAAACCCCATCTCTACCAAAAATACAAAATATAGTCAGGTGGTGGCACTTCGCT	64293		
Qy	240	ATAGTCCCACTACACAGAGGCTGGGGCAGAGAATCACTTGAACCTGGAGGTGGAGG	299		
Db	64292	ATAATCCCACTACTCAGGAGGCTGAGGGAGGAGATCACTTGAACCCAGAGGTGGAGG	64233		
Qy	300	TTGCAGTGAGCCCAAGATTGACCACTGCACCTCAGCTCGGAAATAA 345			
Db	64232	TTGCAGTGAGCCCAAGATCGGCCCACTGCACCTCAGCTCAGCTCGGAAATAA 64187			

RESULT 8
AC026470/c
LOCUS Homo sapiens chromosome 16 clone RP11-44110, complete sequence.
DEFINITION AC026470
ACCESSION AC026470
VERSION AC026470.6 GI:38502331
KEYWORDS HTG.
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

REFERENCE AUTHORS	Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini; Hominidae; Homo.		AX386112 AX386112.1 GI:19579242	
	1 (bases 1 to 169566) DOE Joint Genome Institute, Stanford Human Genome Center and Los Alamos National Laboratory.		Homo sapiens (human)	
	Direct Submission Unpublished		Homo sapiens	
	2 (bases 1 to 169566) DOE Joint Genome Institute. Direct Submission		Homo sapiens	
REFERENCE AUTHORS	3 (bases 1 to 169566) DOE Joint Genome Institute. Direct Submission		Homo sapiens	
	4 (bases 1 to 169566) Stanford Human Genome Center and Los Alamos National Laboratory. DOE Joint Genome Institute		Homo sapiens	
	Submitted (25-NOV-2003) DOE Joint Genome Institute, 2800 Mitchell Drive, Walnut Creek, CA 94598, USA		Homo sapiens	
	On Nov 25, 2003 this sequence version replaced gi:19072800. Draft Sequence Produced by DOE Joint Genome Institute		Homo sapiens	
REFERENCE AUTHORS	www.jgi.doe.gov Finishing Completed at Stanford Human Genome Center and Los Alamos National Laboratory		Homo sapiens	
	www.ehgc.stanford.edu		Homo sapiens	
	Quality: Phrap Quality >=40 99.9% of Sequence; Estimated Total Number of Errors is 0.1.		Homo sapiens	
	Location/Qualifiers		Homo sapiens	
FEATURES source	1.169566		Homo sapiens	
	/organism="Homo sapiens"		Homo sapiens	
	/mol_type="genomic DNA"		Homo sapiens	
	/db_xref="taxon:9606"		Homo sapiens	
ORIGIN	/chromosomes="16"		Homo sapiens	
	/clone="RP11-44110"		Homo sapiens	
	Query Match 43.3%; Score 216.4; DB 8; Length 169566;		Homo sapiens	
	Best Local Similarity 77.5%; Pred. No. 2.5e-55;		Homo sapiens	
QY	Matches 262; Conservative 0; Mismatches 76; Indels 0; Gaps 0;		Homo sapiens	
	30 TAGCATATGGACCTGGGATGGAGTTAGCGCTCAGTTAATAGTAATCAATAGCAGGTG 89		Homo sapiens	
	Db 84067 TAGCCTGGGCAACAGAGTCAGACTTCATCTATGAAATTTAAATTTGGCCAGCG 84008		Homo sapiens	
	QY 90 CGGTGGCTCATGCTGTATTTCCAGCATTGGGAGACCGAGTTGGTGGATCACTTGAG 149		Homo sapiens	
DB	84007 TGGTGGCTCATGCTGTATTTCCAGCATTGGGAGACCGAGTTGGTGGATCACTTGAG 83948		Homo sapiens	
	QY 150 AGCAGAGTTTGAACACGAGCTGGCCAAACATGGCAAAACATCTCTTAATAAAATACA 209		Homo sapiens	
	Db 83947 GTCCAGAGTTTGAACACGAGCTGGCCAAACATGGCAAAACATCTCTTAATAAAATACA 83888		Homo sapiens	
	QY 210 AAAATTAGCAGGTGGTGGGCTGGGCTGGGCTGGGCTGGGCTGGGCTGGGCTGGGCTGG 269		Homo sapiens	
DB	83887 AAAATTAGCAGGTGGTGGGCTGGGCTGGGCTGGGCTGGGCTGGGCTGGGCTGGGCTGG 83828		Homo sapiens	
	QY 270 GAAGATCACTTGAACCTGGGAGTGGGAGTGGGAGTGGGAGTGGGAGTGGGAGTGGGAGTGG 329		Homo sapiens	
	Db 83827 GGAGAATCACTTGAACCTGGGAGTGGGAGTGGGAGTGGGAGTGGGAGTGGGAGTGGGAGTGG 83768		Homo sapiens	
	QY 330 TCCAGCTGGGCAAAAGGTAATTAATCACTTACTT 367		Homo sapiens	
DB	83767 TCCAGCTGGGCAAAAGGTAATTAATCACTTACTT 83730		Homo sapiens	
	RESULT 9		Homo sapiens	
	AX386112		Homo sapiens	
	LOCUS		Homo sapiens	
DEFINITION	Sequence 1040 from Patent WO0214500.		Homo sapiens	
	662 bp DNA linear PAT 19-MAR-2002		Homo sapiens	
	AX386112		Homo sapiens	
	Sequence 1040 from Patent WO0214500.		Homo sapiens	

University School of Medicine, 4444 Forest Park Parkway, St. Louis, MO 63108, USA
 4 (bases 1 to 123576)
 Waterston,R.H.
 Direct Submission
 Submitted (18-SEP-2001) Genome Sequencing Center, Washington University School of Medicine, 4444 Forest Park Parkway, St. Louis, MO 63108, USA
 5 (bases 1 to 123576)
 Waterston,R.
 Direct Submission
 Submitted (09-JAN-2002) Department of Genetics, Washington University, 4444 Forest Park Avenue, St. Louis, Missouri 63108, USA
 6 (bases 1 to 123576)
 Wilson,R.K.
 Direct Submission
 Submitted (16-APR-2005) Genome Sequencing Center, Washington University School of Medicine, 4444 Forest Park Parkway, St. Louis, MO 63108, USA
 On Aug 9, 2001 this sequence version replaced gi:14488387.

 Center: Washington University Genome Sequencing Center
 Center code: WUGSC
 Web site: <http://genome.wustl.edu>
 Contact: submissions@watson.wustl.edu

 Summary Statistics
 Center project name: H_NH0218N06

NOTICE:

This sequence was finished as follows unless otherwise noted:
 all regions were double stranded, sequenced with an alternate chemistry, or covered by high quality data (i.e., phred quality >= 30); an attempt was made to resolve all sequencing problems, such as compressions and repeats; all regions were covered by sequence from more than one subclone; and the assembly was confirmed by restriction digest.

MAPPING INFORMATION:

Mapping information for this clone was provided by Dr. Wes Warren, Department of Genetics, Washington University, St. Louis MO. For additional information about the map position of this sequence, see <http://genome.wustl.edu>

SOURCE INFORMATION:

The RP11-11 human BAC library was made from the blood of one male donor, as described by Osoegawa,K., Woon,P.Y., Zhao,B., Frengen,E., Tatenno,M., Catanese,J.J. and de Jong,P.J. (1998) An improved approach for construction of bacterial artificial chromosome libraries. Genomics 51:1-8. The clone may be obtained either from Research Genetics, Inc. (<http://www.resgen.com>) or Pieter de Jong and coworkers at <http://www.chori.org>
 VECTOR: pBACe3.6

NEIGHBORING SEQUENCE INFORMATION:

The clone sequenced to the left is RP11-474G23, 2000 bp overlap. Actual start of this clone is at base position 144971 of RP11-474G23; actual end is at base position 123576 of RP11-218N6.

RP11-218N6 contains a single plasmid region from 42433 to 42444.

FEATURES
source

1. 123576
 Location/Qualifiers
 /organism="Homo sapiens"
 /mol_type="genomic DNA"
 /db_xref="taxon:9606"
 /chromosome="2"
 /clone="RP11-218N6"
 /clone_lib="RP11-11"
 complement(16243..42944)
 /gene="DKFZP566K1924"
 complement(join(16243..17259,40390..40540,42455..42944))
 /gene="DKFZP566K1924"

gene

mRNA

CDS
 complement(join(17095..17259,40390..40540,42455..42633))
 /gene="DKFZP566K1924"
 /note="Homo sapiens DKFZP566K1924 protein (DKFZP566K1924), mRNA ; H_NH0218N06.1"
 This gene was based on gi(24308070)"
 /codon_start=1
 /product="unknown"
 /protein_id="AAK93120.1"
 /db_xref="GI:62702194"
 /translation="MGDLPLGLRLSIALRIOPNDGPVPYKVDGDFGONRTIKLLTGS SYKVEVKIKSTLOVENISGGVLVPLELKSKEPDGRVVTYGTDEGVPTPKSGER QPIQITMPPFDIDIGITETVWQVKFYNYHHRDHCQMGSPFSVIEYCKNETRSLMWVK ESFL"
 26142..26379
 /note=".CpG island (%GC=61.3, o/e=0.90, #CpGs=19)"
 42441..43212
 /note=".CpG island (%GC=68.8, o/e=0.74, #CpGs=75)"
 77919..78140
 /note=".CpG island (%GC=60.4, o/e=0.80, #CpGs=16)"
 88517..120686
 /gene="PLEK"
 join(88517..88626,103561..103716,103956..104137,105775..105866,109735..109919,111620..111724,116395..116478,117340..117409,118913..120686)
 /gene="PLEK"
 join(88585..88626,103561..103716,103956..104137,105775..105866,109735..109919,111620..111724,116395..116478,117340..117409,118913..119049)
 /gene="PLEK"
 /note="Homo sapiens pleckstrin (PLEK), mRNA ; H_NH0218N06.2"
 This gene was based on gi(4505878)"
 /codon_start=1
 /product="unknown"
 /protein_id="AAK93121.1"
 /db_xref="GI:62702195"
 /translation="MEPKIRREGVLYKKSQVFNWTKPMVVLLEDGIEFYKKSQNSP KGMIPKGLSTLSTPQDFGRMFVKITTTQQDDHFFQAAFLBERDAWDRDKKAIKC IEGGOKPARKSTRSIRLPETIDLGALVLSMKDTEKI KEINLEKDKKI FNHCPTGNC VIDLVSNQSVNRNREQGLMIASSILNEGILQPADGMSKSAVDGAENPFLDNPDAFY PPSGPFCEENSDDDLVKEEFGVITIKGCLLKQGHRRNRNWKRFIFUREDPAFLH YYDPAGADEPLGAHLHRLGCVVTSVENSNGRKEENLFEIITADEVHYFLQATPKE RTEWIRAIQWASRTGK"

ORIGIN

Query Match 43.0%; Score 214.8; DB 8; Length 123576;
 Best Local Similarity 83.8%; Pred. No. 8e-55;
 Matches 243; Conservative 0; Mismatches 47; Indels 0; Gaps 0;
 QY 63 AGTTAATAGTAACCTCATTAGCAGGTGCGGTGCGTCTATGTCTGTATCCAGCACTTTGG 122
 DB 33261 AGTAAGAGTTTGGATTGCCAGGACGGTGGCTACTCTCTGTATCCAGCACTTTGG 33320
 QY 123 GAGACCGAGTTGGTGGATCACTTGAGAGCAGAGGTTTGAACACAGCTCGCCAACTGG 182
 DB 33321 GAGCCGAGCAGGTGGATCACTTGAGTTCAGAGTTCGAGACAGCGTGGCCAACTGG 33380
 QY 183 CAAACACTATCTCTAATAAATAACAAAATTAGCCAGGTGTGGTGGCACTTGCCTATA 242
 DB 33381 TGAACCCCTCTCTCTAATAAATAACAAAATTAGCCAGGTGTGGTGGCACTTGTGTA 33440
 QY 243 GTCCAGCTACACAGGAGGTGGGGCAGAGAATCACTTCAACCTGGAGGTGGAGGTG 302
 DB 33441 GCCCCAGCTACTCTGGCGGCTGAGGCGAGGAATCGCTTGAACCTGGAGGTGGAGGTG 33500
 QY 303 CAGTGAGCCAAAGATTGCACCACTGCATCTCCAGCTCGAAAAAAGGTAA 352
 DB 33501 CAGTGAGCTGAGATCGCACCTGCATTCAGCTTCGGCAACCAAAAGTGA 33550

RESULT 11

AC146080

LOCUS

AC146080 190306 bp DNA linear PRI 19-MAY-2004

```
DEFINITION Pan troglodytes BAC clone RP43-51K11 from 7, complete sequence.
ACCESSION AC146080
VERSION AC146080.4 GI:47498253
KEYWORDS HTG.
SOURCE Pan troglodytes (chimpanzee)
ORGANISM Pan troglodytes
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini;
Hominoidea; Pan.
1 (bases 1 to 190306)
Tomlinson,C. and Bielicki,L.
The sequence of Pan troglodytes BAC clone RP43-51K11
Unpublished (2001)
2 (bases 1 to 190306)
Wilson,R.K.
Direct Submission
Submitted (01-AUG-2003) Genetics, Genome Sequencing Center, 4444
Forest Park Parkway, St. Louis, MO 63108, USA
3 (bases 1 to 190306)
Wilson,R.K.
Direct Submission
Submitted (24-APR-2004) Genetics, Genome Sequencing Center, 4444
Forest Park Parkway, St. Louis, MO 63108, USA
4 (bases 1 to 190306)
Wilson,R.K.
Direct Submission
Submitted (19-MAY-2004) Washington University School of Medicine,
Genome Sequencing Center, 4444 Forest Park Parkway, St. Louis, MO
63108, USA
On May 19, 2004 this sequence version replaced gi:46559537.
----- Genome Center
Center: Washington University Genome Sequencing Center
Center code: WUGSC
Web site: http://genome.wustl.edu
Contact: submission@wustl.wustl.edu
----- Summary Statistics
-----
Center project name: C_PT051K11
-----
NOTICE:
This sequence was finished as follows unless otherwise noted:
all regions were double stranded, sequenced with an alternate
chemistry, or covered by high quality data (i.e., phred quality >=
30); an attempt was made to resolve all sequencing problems, such
as compressions and repeats; all regions were covered by sequence
from more than one subclone; and the assembly was confirmed by
restriction digest.
MAPPING INFORMATION:
Mapping information for this clone was provided by Dr. Wes Warren,
Department of Genetics, Washington University, St. Louis MO. For
additional information about the map position of this sequence, see
http://genome.wustl.edu
SOURCE INFORMATION:
The RPCI-43 BAC Library has been constructed by Chung-Li Shu. DNA
was isolated from white blood cells obtained from a male chimpanzee
(Pan troglodytes, Clint', Yerkes #C0471; birthdate: 6-6-80). The
clone and detailed information can be obtained from ResGen
(http://www.resgen.com) or Pieter de Jong and co-workers at
http://www.bacpac.chori.org.
NEIGHBORING SEQUENCE INFORMATION:
This sequence is the entire insert of the clone.
Location/Qualifiers
1. .190306
/organisms="Pan troglodytes"
/mol_type="genomic DNA"
/db_xref="taxon:9598"
/chromosome="7"
/map="7"
/clones="RP43-51K11"
```

```
ORIGIN
/clone_lib="RPCI-43"
Query Match 43.0%; Score 214.8; DB 8; Length 190306;
Best Local Similarity 82.6%; Pred. No. 7.9e-55;
Matches 246; Conservative 0; Mismatches 52; Indels 0; Gaps 0;
Qy 67 AATAGTAACCTCATTAGCCAGGTGGGTGGCTCATGTCTGATTCCAGCAGCATTGGGAGA 126
Db 133566 AATAAGAAAAAGAGGCCAGGTGAGGTGGCTCATGCTGTATATCCAGCAGCATTGGGAGG 133625
Qy 127 CCAGATTGGGTGGATCATTCTTGAGAGCAGGAGTGTGACACAGCAGCTGGCCAAACATGGCAAA 186
Db 133626 CTGAGGTGGGTGGATCATTGAGTTCAGGATTCAGACACAGCAGCTGGCCAAAGATGTGAA 133685
Qy 187 ACTATATCTCTAATAAAAAATACAAAAATTAGCCAGGTGTGGTGGCAGCTTGCCTATAGTCC 246
Db 133686 ACCCATCTCTGTAAAAAATACAAAAATTAGCCAGGTGTGGTGGCAGCTGCTGTAAATCC 133745
Qy 247 CAGCTACAGGAGGCTGGGGCAGAGCAATCATTGAACCTGGGAGGTGGAGGTGGCAGT 306
Db 133746 CAGCTACTCAGGAGGCTGGGGCAGGAGAACTCTTGAACCTGGGAGGAGGTGGCAGT 133805
Qy 307 GAGCCAAAGATTGCACCACTGCACCTCCAGCCTGGAAAAAGGGTAATTAATAACTTTA 364
Db 133806 GAGCCGAGATTGCACCACTGCACCTCCAGCCTGGGTGGACAGAGTGAGACTGTGACTTAA 133863
RESULT 12
AP002789/c AP002789 90000 bp DNA linear PRI 15-MAR-2003
LOCUS Homo sapiens genomic DNA, chromosome 11q, clone:RP11-716L7,
DEFINITION complete sequence.
ACCESSION AP002789
VERSION AP002789.3 GI:15320481
KEYWORDS HTG.
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini;
Hominoidea; Homo.
1
Hattori,M., Ishii,K., Toyoda,A., Taylor,T.D., Hong-Seog,P.,
Fujiyama,A., Yada,T., Totoki,Y., Watanabe,H. and Sakaki,Y.
TITLE Homo sapiens genomic DNA
JOURNAL Published Only in Database (2000)
REFERENCE 2 (bases 1 to 90000)
Hattori,M., Ishii,K., Toyoda,A., Taylor,T.D., Hong-Seog,P.,
Fujiyama,A., Yada,T., Totoki,Y., Watanabe,H. and Sakaki,Y.
Direct Submission
Submitted (11-JUL-2000) Masahira Hattori, The Institute of Physical
and Chemical Research (RIKEN), Genomic Sciences Center (GSC),
1-7-22 Suehiro-chou,Tsukumi-ku, Yokohama, Kanagawa 230-0045, Japan
(E-mail:hattori@gs.c.riken.go.jp, URL:http://hgp.gsc.riken.go.jp/,
Tel:81-45-503-9111, Fax:81-45-503-9170)
COMMENT On Aug 27, 2001 this sequence version replaced gi:11071948.
FEATURES
Location/Qualifiers
1. .90000
/organism="Homo sapiens"
/mol_type="genomic DNA"
/db_xref="taxon:9606"
/chromosome="11"
/map="11q"
/clone="RP11-716L7"
ORIGIN
Query Match 43.0%; Score 214.6; DB 8; Length 90000;
Best Local Similarity 79.8%; Pred. No. 9.3e-55;
Matches 253; Conservative 0; Mismatches 64; Indels 0; Gaps 0;
Qy 69 TAGTAACCTCATTAGCCAGGTGGGTGGCTCATGTCTGATTCCAGCAGCATTGGGAGACC 128
Db 70289 TAAGAGCTTCTGCACCCAGGCGGTGGCTCAGCCCTGTATATCCAGCAGCTTTGGGAGGCC 70230
```

QY 129 GAGTTGGGTGATCACTTGAGAGCAGGAGTTTGAGACCAGCTGCGCAACATGGCAAAAC 188
 |||||
 Db 70229 GAGGTGGGTGGATCACTCAGCTCAGGAGTTTGAGACCAGCTGCGCAACATGGTGAAC 70170
 |||||
 QY 189 ACTATCTCTATAAAATACAAAATAGCCAGGTGTGGTGGCACTTGCCTATAGTCCCA 248
 |||||
 Db 70169 CTGTCTCTACTAAATAATACAAAATAGCCAGGCGTGTGGCACATCCCTGTAGTCCCA 70110
 |||||
 QY 249 GCTACACAGGAGCTGGGGCAGAGAATCACTTGAACCTGGAGGTGGAGGTGGCAGTGA 308
 |||||
 Db 70109 GCTACTCAGGAGCTGAGCAGAGAATAGCTTGAACCCGGAGGTGGAGGTGGCAGCA 70050
 |||||
 QY 309 GCCAAGATTGCACCACTGCATCTCCAGCTCGAAAAAGGGTAATTAATACTTTACTTTG 368
 |||||
 Db 70049 GCCAAGATTATACCACTAACTCCAGCTGGCGCACTGTGTGAGACTCCATCTCAAAAA 69990
 |||||
 QY 369 CAACCATAGCTGCTTCT 385
 |||||
 Db 69989 AAAAAAAGAGCTTCT 69973
 |||||

RESULT 13
 BV632439/c
 LOCUS
 DEFINITION S215P60370FH7.T0 Clara Pan troglodytes troglodytes STS 15-APR-2005
 sequence tagged site.
 ACCESSION BV632439
 VERSION BV632439.1 GI:62622617
 KEYWORDS STS.
 SOURCE Pan troglodytes troglodytes
 ORGANISM Pan troglodytes troglodytes
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini;
 Hominidae; Pan.
 1 (bases 1 to 726)
 Mikkelson, T.S., Hillier, W.L., Eichler, E.E., Zody, M.C. and
 Jaffe, D.B.
 Initial Sequence of the Chimpanzee Genome and Comparison with the
 Human Genome
 Unpublished (2005)

Contact: Michael C. Zody
 Broad Institute of MIT and Harvard
 320 Charles Street, Cambridge, MA 02141, USA
 Tel: 6172580933
 Fax: 6172580903
 Email: mczody@broad.mit.edu
 Primer A: No sequence submitted
 Primer B: No sequence submitted
 STS size: 726
 Protocol:
 23,021,928 chimpanzee whole genome shotgun reads were aligned to
 the Human genome NCBI
 Build 34 (hg16, July 2003). Chimp WGS reads were from 9 donors,
 including Clint (Pan
 troglodytes verus), 3 other Pan troglodytes verus chimps
 (Donald, Karlien, Yvonne), 3 Pan
 troglodytes troglodytes chimps (Noemie, Masuku, Clara) and 2 chimps
 of unknown origin
 (Gon, Unknown Chimp). Common names: Pan troglodytes verus is the
 western chimp and Pan
 troglodytes troglodytes is the central chimp. To be included in
 chimpanzee SNP discovery, a
 read must be at least 500bp in length, at least 50% of its base
 calls must have Phred
 score >= 20, at least 30% of its base calls must satisfy
 SNOS(30,25) (single strand NOS, the
 base in question has Phred score >= 30, the surrounding 10 bases in
 the read have Phred
 score >= 25), and the read must have at least 200 bp SNOS(30,25)
 bases. Reads not uniquely
 placed in the genome and read pairs whose two ends were not

consistently placed were
 discarded. After above filtering, NOS(30,25) standard was applied
 to all pairs of
 overlapping reads to call NOS bases and SNPs. Alignments (between
 two reads) with less
 than 100 NOS bases or with SNP rate > 0.01 were discarded. To
 exclude alignment between two
 copies of a single read, comparisons between two reads that share
 95% of their genome
 alignments (>=95% bases of read A and >=95% bases of read B were
 placed at the same locus
 of human genome) were discarded.

FEATURES

source
 1..726
 /organism="Pan troglodytes troglodytes"
 /mol_type="genomic DNA"
 /sub_species="troglodytes"
 /db_xref="taxon:37011"
 /clone_lib="Clara"
 <1..>726

STS
ORIGIN

Query Match 42.9%; Score 214.4; DB 10; Length 726;
 Best Local Similarity 80.4%; Pred. No. 1.4e-54;
 Matches 251; Conservative 0; Mismatches 61; Indels 0; Gaps 0;
 QY 46 GGATGGAGTTAGCGCTCAGTTAATAGTAATCTATTAGCCAGGTGCGGTGCTCATGTCGT 105
 |||||
 Db 312 GAATTAATTAGTGCCCTTAAAGAGACCCAGAGCCAGGAGTGTGTGCTCACACCTG 253
 |||||
 QY 106 TATTCACGACATTTGGGAGACCGAGTTGGTGGATCACTTTGAGAGCAGGAGTTTGAGAC 165
 |||||
 Db 252 TAATCCAGCACATTTGGAGAGCCGAGGTGGCGAGATCACTTTGAGTCAGGAGTTTGAGAC 193
 |||||
 QY 166 CAGCTGGCCCAATGCGCAAAACATCTATCTCTAATAAATAACAAAAATAGCCAGGTGT 225
 |||||
 Db 192 TAGCCTGGCCCAATGCGCAAAACCTATTCTTCTAATAAATAACAAAAATAGCCAGATGT 133
 |||||
 QY 226 GGTGGCACTTGCCTATAGTCCAGCTACACAGAGCTGGGGCAGAGAGATCACTTTGAAC 285
 |||||
 Db 132 GGTGGCACATGCTCTTAATTCAGCTACTCAGGAGCTGAGGATGAGAAATTTCTTGAAC 73
 |||||
 QY 286 CTGGGAGGTGGAGGTTCAGTGAGCCAAAGATTGCACCATCTGCACCTCAGCCTGGAAAAA 345
 |||||
 Db 72 CCAGGAGATGAGGTTCAGGAGAGCCAGAGATGACACCTGCACCTCAGCCTGGGTGATA 13
 |||||
 QY 346 AGGGTAATTAAT 357
 |||||
 Db 12 AAGCAAGAGAAAT 1
 |||||

RESULT 14

AC114889/c

LOCUS

DEFINITION

AC114889

VERSION

KEYWORDS

SOURCE

ORGANISM

REFERENCE

AUTHORS

1 (bases 1 to 193369)

Akher, N., Antonellis, A., Ayele, K., Beckstrom-Sternberg, S.M.,

Benjamin, B., Blakesley, R.W., Bouffard, G.G., Brinkley, C., Brooks, S.,

Carlaga, K., Coleman, B., Engle, J., Granite, S., Guan, X., Gupta, J.,

Haghighi, P., Han, J., Hansen, N., Ho, S.-L., Idol, J.R., Karlins, E.,

Laric, P., Lee-Lin, S.-O., Legaspi, R., Maduro, Q.L., Maduro, V.B.,

Margulies, E.H., Masiello, C., Maskeri, B., McDevell, J.,

Paguirigan, C., Pearson, R., Portnoy, M.E., Prasad, A.,

Reddix-Dugue, N., Schandler, K., Schueler, M.G., Sison, C.,

Sison, C.,

Sison, C.,

Sison, C.,

Sison, C.,

Sison, C.,

Sison, C.,

Stantripop,S., Thomas,J.W., Thomas,P.J., Touchman,J.W., Vogt,J.L.,
Wetherby,K.D.; Wiggins,L., Young,A. and Green,E.D.
NISC Comparative Sequencing Initiative
Unpublished
2 (bases 1 to 193369)
Green,E.D.
Direct Submission
Submitted (13-MAR-2002) NIH Intramural Sequencing Center, 8717
Government Circle, Gaithersburg, MD 20877, USA
3 (bases 1 to 193369)
Green,E.D.
Direct Submission
Submitted (18-JAN-2003) NIH Intramural Sequencing Center, 8717
Government Circle, Gaithersburg, MD 20877, USA
On Jan 18, 2003 this sequence version replaced gi:21622727.
----- Genome Center
Center: NIH Intramural Sequencing Center
Center code: NISC
Web site: http://www.nisc.nih.gov
Contact: nisc_zoo@nhgri.nih.gov
----- Project Information
Center project name: cld
Center clone name: 064L14

The sequence data in this record represents an 'enhanced'
version of a Phase 2 submission. Specifically, the indicated
order and orientation of each sequence contig has been
established using one or more of the following: read-pair
data from individual subclones, overlaps with neighboring
clones, alignment with available reference sequence (e.g.,
human), and/or confirmation by PCR testing. In addition,
the sequence assembly is based on at least 8x average
coverage in Q20 bases and has been reviewed to rule out
gross misassemblies, the low-quality ends of sequence
contigs have been trimmed away, and each base is associated
with a Phrap-derived quality score.

----- Summary Statistics
Sequencing vector: plasmid; n/a; 100% of reads
Chemistry: Dye-terminator Big Dye; 100% of reads
Assembly program: Phrap; version 0.990319
Consensus quality: 192665 bases at least Q40
Consensus quality: 192877 bases at least Q30
Consensus quality: 192937 bases at least Q20
Insert size: 120000; agarose-fp
Insert size: 192969; sum-of-contigs
Quality coverage: 15.56x in Q20 bases; agarose-fp
Quality coverage: 9.68x in Q20 bases; sum-of-contigs

* NOTE: This is a 'working draft' sequence. It currently
* consists of 5 contigs. Gaps between the contigs
* are represented as runs of N. The order of the pieces
* is believed to be correct as given, however the sizes
* of the gaps between them are based on estimates that have
* provided by the submittor.
* This sequence will be replaced
* by the finished sequence as soon as it is available and
* the accession number will be preserved.

1 2008: contig of 2008 bp in length
2009 2108: gap of unknown length
2109 25979: contig of 23871 bp in length
25980 26079: gap of unknown length
26080 158388: contig of 132309 bp in length
158389 158488: gap of unknown length
158489 188878: contig of 30390 bp in length
188879 188978: gap of unknown length
188979 193369: contig of 4391 bp in length.

Location/Qualifiers
1. .193369
/organism="Pan troglodytes"
/mol_type="genomic DNA"
/db_xref="taxon:9598"
/clone="RP43-64L14"
/clone_lib="RP43"

FEATURES
source

misc_feature 1. .63570
/notes="clone overlaps with GenBank Accession Number
AC116439 clone RP43-88G1 (center project name c1c)"
1. .2008
/notes="assembly_fragment
missing T7 clone end on 5' end of insert"
2009. .2108
/estimated length=unknown
2109. .25979
/notes="assembly_fragment"
25980. .26079
/estimated length=unknown
26080. .158388
/notes="assembly_fragment"
158389. .158488
/estimated length=unknown
158489. .188878
/notes="assembly_fragment"
161692. .193369
/notes="clone overlaps with GenBank Accession Number
AC114888 clone RP43-159M22 (center project name c1c)"
188879. .188978
/estimated length=unknown
188979. .193369
/notes="assembly_fragment
clone end:SP6
vector_side:right"
ORIGIN
Query Match 42.9%; Score 214.4; DB 14; Length 193369;
Best Local Similarity 84.0%; Pred. No. 1e-54;
Matches 242; Conservative 0; Mismatches 46; Indels 0; Gaps 0;
QY 65 TTAATAGTAATCTCAATAGCCAGGTGCGGTGCTCATGCTGTATCCAGCACATTTGGGA 124
Db 121917 TTAATAATATTCAAAACAGCCAGGTGCGGTGCTCATGCTGTATCCAGCACATTTGGGA 121858
QY 125 GACCGAGTTGGTGGATCATTGAGAGCAGGATTTGAGACAGCTGCGCCAAACATGGCA 184
Db 121857 GACCAAGCGGGCAGATCACTGAAGTCAGAGATTGAGACAGCTGCGCCAAACATGGTG 121798
QY 185 AAACACTATCTCTAATAAAAAATACAAAAATTAGCCAGGTGTTGGTGCACTTGCCTATAGT 244
Db 121797 AAACCTGTCTCTACTAAAAATACAAAAATTAGCTGGGTGGTGACATGCTGTAAAT 121738
QY 245 CCCAGCTACACAGAGGCTGGGGCAGAAGATCACTTGAACCTGGGAGGTGGAGTTGCA 304
Db 121737 CCCAGCTACTTAGGAGGCTGAGCGAGGAGAAATCTTGAACCTGGGAGGTGGAGTTGCA 121678
QY 305 GTGAGCCAGATTGCACCACCTGCACCTCCAGCTGGAAAAAAGGGTAA 352
Db 121677 GTGAGCCAGATTGTGCCATTGTACTCCAGCTTGGCCAAACAGAGCAA 121630

RESULT 15

AC074389 AC074389 195782 bp DNA linear PRI 08-OCT-2003
LOCUS Homo sapiens BAC clone RP11-510K8 from 7, complete sequence.
DEFINITION AC074389
ACCESSION AC074389
VERSION AC074389.8 GI:18042461
KEYWORDS HTG.
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini;
Hominiidae; Homo.
REFERENCE 1 (bases 1 to 195782)
AUTHORS Sulston,J.E. and Wilson,R.
TITLE Toward a complete human genome sequence
JOURNAL Genome Res. 8 (11), 1097-1108 (1998)
PUBMED 9847074
REFERENCE 2 (bases 1 to 195782)
AUTHORS Cordes,M., Doebber,A., Hawkins,M. and Kozlowski,A.

TITLE
JOURNAL
REFERENCE
3 (bases 1 to 195782)
Waterston,R.H.
TITLE
JOURNAL
Submitted (30-JUL-2000) Genome Sequencing Center, Washington
University School of Medicine, 4444 Forest Park Parkway, St. Louis,
MO 63108, USA
4 (bases 1 to 195782)
Waterston,R.H.
TITLE
JOURNAL
Submitted (03-JAN-2002) Genome Sequencing Center, Washington
University School of Medicine, 4444 Forest Park Parkway, St. Louis,
MO 63108, USA
5 (bases 1 to 195782)
Waterston,R.
TITLE
JOURNAL
Submitted (10-JAN-2002) Department of Genetics, Washington
University, 4444 Forest Park Avenue, St. Louis, Missouri 63108, USA
6 (bases 1 to 195782)
Waterston,R.
TITLE
JOURNAL
Submitted (29-APR-2003) Department of Genetics, Washington
University, 4444 Forest Park Avenue, St. Louis, Missouri 63108, USA
7 (bases 1 to 195782)
Wilson,R.
TITLE
JOURNAL
Submitted (08-OCT-2003) Department of Genetics, Washington
University, 4444 Forest Park Avenue, St. Louis, Missouri 63108, USA
On Jan 3, 2002 this sequence version replaced gi:14209801.
----- Genome Center
Center: Washington University Genome Sequencing Center
Center code: WUGSC
Web site: <http://genome.wustl.edu>
Contact: sapiens@watson.wustl.edu
----- Summary Statistics

Center project name: H_NH0510K08

NOTICE: This sequence may not represent the entire insert of this
clone. It may be shorter because we only sequence overlapping
clone sections once, or longer because we provide a small overlap
between neighboring data submissions.

This sequence was finished as follows unless otherwise noted:
all regions were double stranded, sequenced with an alternate
chemistry, or covered by high quality data (i.e., phred quality >= 30); an attempt was made to resolve all sequencing problems, such
as compressions and repeats; all regions were covered by sequence
from more than one subclone; and the assembly was confirmed by
restriction digest.

MAPPING INFORMATION:

The sequence of this clone was established as part of a mapping and
sequencing collaboration between the NHGRI Chromosome 7 Mapping
Project (Eric D. Green, Director), John D. McPherson in the
Department of Genetics (Washington University), and the Washington
University Genome Sequencing Center. For additional information
about the map position of this sequence, see
<http://www.nhgri.nih.gov/DIR/OTB/CHR7>, send
mailto:egreen@nhgri.nih.gov , or see <http://genome.wustl.edu>

SOURCE INFORMATION:

The RPCI-11 human BAC library was made from the blood of one male
donor, as described by Osoegawa,K., Woon,P.Y., Zhao,B., Frenken,E.,
Tateno,M., Catanese,J.J. and de Jong,P.J. (1998) An improved
approach for construction of bacterial artificial chromosome
libraries. Genomics 51:1-8. The clone may be obtained either from
Research Genetics, Inc. (<http://www.reagen.com>) or Pieter de Jong
and coworkers at <http://www.chori.org>
VECTOR: pBACe3.6

NEIGHBORING SEQUENCE INFORMATION:

Data from AC093734 was used to finish this clone, AC074389.

The clone sequenced to the left is RP11-16P10, 2000 bp overlap.
Actual start of this clone is at base position 112086 of RP11-16P10
actual end is at base position 195782 of RP11-510K8.

FEATURES	Location/Qualifiers
source	1..195782
	/organism="Homo sapiens"
	/mol_type="genomic DNA"
	/db_xref="taxon:9606"
	/chromosome="7"
	/map="7"
	/clone="RP11-510K8"
	/clone_lib="RPCI-11"
repeat_region	3..269
	/rpt_family="L1"
repeat_region	270..372
	/rpt_family="MIR"
repeat_region	558..736
	/rpt_family="Alu"
repeat_region	807..844
	/rpt_family="(TTCA)n"
repeat_region	946..1126
	/rpt_family="C-rich"
repeat_region	1099..1278
	/rpt_family="CT-rich"
repeat_region	1417..1471
	/rpt_family="CT-rich"
repeat_region	1485..1595
	/rpt_family="L1"
repeat_region	1621..1647
	/rpt_family="(CA)n"
repeat_region	2190..2260
	/rpt_family="(GGCTG)n"
repeat_region	2261..2383
	/rpt_family="MIR"
repeat_region	2804..3125
	/rpt_family="L2"
repeat_region	3175..3277
	/rpt_family="MaLR"
repeat_region	3599..3761
	/rpt_family="L2"
repeat_region	3762..4059
	/rpt_family="Alu"
repeat_region	4060..4157
	/rpt_family="L2"
repeat_region	4173..4432
	/rpt_family="L2"
repeat_region	5572..5740
	/rpt_family="MIR"
repeat_region	5756..6058
	/rpt_family="Alu"
repeat_region	6506..6651
	/rpt_family="L2"
repeat_region	6652..6934
	/rpt_family="Alu"
repeat_region	6935..7090
	/rpt_family="L2"
repeat_region	7091..7404
	/rpt_family="Alu"
repeat_region	7503..7785
	/rpt_family="Alu"
repeat_region	7787..7815
	/rpt_family="ERV1"
repeat_region	7816..8151
	/rpt_family="Alu"
repeat_region	8152..8534
	/rpt_family="ERV1"
repeat_region	8535..8804
	/rpt_family="Alu"
repeat_region	10374..10668
	/rpt_family="Alu"
repeat_region	10854..11031

```
repeat_region /rpt_family="MER1_type"
11051. .11212
/rpt_family="ERV1"
repeat_region 11235. .11328
/rpt_family="MIR"
repeat_region 11692. .11800
/rpt_family="MaLR"
repeat_region 11801. .12113
/rpt_family="Alu"
repeat_region 12114. .12415
/rpt_family="MaLR"
repeat_region 12559. .13339
/rpt_family="(TG)n"
repeat_region 13357. .13530
/rpt_family="(TG)n"
repeat_region 13564. .13738
/rpt_family="(TG)n"
misc_feature 13740. .14544
/note="CpG island (%GC=69.6, o/e=0.82, #CpGs=65)"
repeat_region 14474. .14593
/rpt_family="MaLR"
repeat_region 14597. .14909
/rpt_family="Alu"
repeat_region 14920. .15038
/rpt_family="L1"
misc_feature 15062. .15480
/note="CpG island (%GC=65.9, o/e=0.79, #CpGs=33)"
repeat_region 15536. .15708
/rpt_family="L1"
repeat_region 16108. .16168
/rpt_family="Alu"
repeat_region 16169. .16379
/rpt_family="Alu"
repeat_region 16380. .16429
/rpt_family="Alu"
repeat_region 16430. .16732
/rpt_family="Alu"
```

Query Match 42.9%; Score 214.4; DB 8; Length 195782;
Best Local Similarity 84.0%; Pred. No. 1e-54;
Matches 242; Conservative 0; Mismatches 46; Indels 0; Gaps 0;

Qy	65	TTAATAGTAAC	TATTAGCCAGGTGGGTGGCTCATGTCTGTATTC	CCAGCAC	TTTGGGA	124
Db	55184	TTTATATTGAGTCCCGCGGGTGCAGTGGCTCACTCCTGTAGTCC	CAGCAC	TTTGGGA	55243	
Qy	125	GACCGAGTTGGGTGGATCACTTGAGAGCAGGAGTTTGAGACCAGCCTGGCCAA	CATGGCA	184		
Db	55244	GGCTGAGCGGGTGGATCACCTGAGATCAGGAGTTCGAGACCAGCCTGGCCAA	CATGGTG	55303		
Qy	185	AAACACTATCTCTAATAAAAAATACAAAAATTAGCCAGGTGTGGTGGCACTTGCCTATAGT	244			
Db	55304	AAACCCCGTCTCTATCTAATAAATAAATAAATAAATTAGCCAGGTGTGGTGGCAGTGCCTATAAT	55363			
Qy	245	CCGAGCTACACAGGAGGCTGGGCGAGAGAATCACTTGAACCTGGAGGTGGAGTTGCA	304			
Db	55364	CCTAGCTACTCGGGAGCGCGAGGAGAGATCACTTGAATCCAGGAGGTGGAGTTGCA	55423			
Qy	305	GTGAGCCAGATTGACACCTGCACCTCCAGCCTGGAAAAAAGGGTAA	352			
Db	55424	GTGAGCCGAGATCGCACCTGCACCTCCAGCCTGGGGGCAAGAGTGA	55471			

GenCore version 5.1.7
Copyright (c) 1993 - 2006 Bioceleration Ltd.

OM nucleic - nucleic search, using sw model

Run on: February 9, 2006, 13:16:14 ; Search time 344.27 Seconds
(without alignments)
9679.442 Million cell updates/sec

Title: US-10-607-806-1_COPY_1_500

Perfect score: 499.6

Sequence: 1 gacctacctcgacctttgtg.....accacataagaggactga 500

Scoring table: IDENTITY_NUC

Gapop 10_0 , Gapext 1.0

Searched: 4996997 seqs, 332346308 residues

Total number of hits satisfying chosen parameters: 9993994

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : N_Geneseq_21.*

1: Geneseqn1980s.*
2: Geneseqn1990s.*
3: Geneseqn2000s.*
4: Geneseqn2001as.*
5: Geneseqn2001bs.*
6: Geneseqn2002as.*
7: Geneseqn2002bs.*
8: Geneseqn2003as.*
9: Geneseqn2003bs.*
10: Geneseqn2003cs.*
11: Geneseqn2003ds.*
12: Geneseqn2004as.*
13: Geneseqn2004bs.*
14: Geneseqn2005s.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	499.6	100.0	12174	12 ADI35082	ADI35082 Human PLA
2	499.6	100.0	12174	12 ADJ09983	Adj09983 Human pho
3	214.8	43.0	662	6 ABN61073	Abn61073 Human can
4	214.8	43.0	662	14 ACL58180	ACL58180 Human col
5	214.8	43.0	35804	14 ADZ59751	Adz59751 Human CB1
6	214.8	42.8	596	5 ABAL7596	Abal7596 Human ner
7	213.4	42.7	587	6 ABN64811	Abn64811 Human can
8	213.4	42.7	587	14 ACL57703	ACL57703 Human col
9	213.4	42.7	9469	4 AAK79514	AAK79514 Human imm
10	213.4	42.7	168174	6 ABT11173	ABT11173 Human 5-1
11	213.4	42.7	168273	6 ABT11114	ABT11114 Human 5-1
12	212.6	42.6	3660	12 ADQ63479	Adq63479 Novel hum
13	211.8	42.4	10252	4 AAK90931	Aak90931 Human dig
14	211.8	42.4	10252	5 AAS31966	Aas31966 Human liv
15	211.8	42.4	10252	6 ABN90321	Abn90321 Human liv
16	211.8	42.4	10252	11 ADJ15234	Adj15234 Human liv
17	211.6	42.4	127722	12 ADQ97301	Adq97301 Human can
18	211.4	42.3	91080	6 ABX08336_16	Continuation (17 o
19	211.4	42.3	91134	12 ADO50281_16	Continuation (17 o

20	211.4	42.3	91138	12 ADN97989_16	Continuation (17 o
21	211.4	42.3	91139	12 ADJ25985_16	Continuation (17 o
22	211.4	42.3	91140	14 AEB85185_16	Continuation (17 o
c 23	211	42.2	3723	11 ADM01799	Adm01799 Human CDN
24	211	42.2	11337	4 AAL38286	Aal38286 Human mus
25	211	42.2	11337	8 ABX59274	Abx59274 CDNA enco
26	211	42.2	11337	12 ADJ30024	Adj30024 Human mus
c 27	210.6	42.2	4622	4 ABK42254	Abk42254 Genomic s
c 28	210.6	42.2	4622	9 ADB60410	Adb60410 Connectiv
29	210.6	42.2	11813	12 ADM80714	Adm80714 Human cla
30	210.6	42.2	52216	4 AAH28355	Aah28355 Nucleotid
31	210.6	42.2	52216	6 ABL50307	AbL50307 Human mus
c 32	210.4	42.1	10113	4 AAS32838	Aas32838 Human gen
c 33	210.4	42.1	11655	4 AAS32839	Aas32839 Human gen
c 34	210.4	42.1	30652	14 AEB94744	Aeb94744 Human eif
35	210	42.0	79544	12 ADQ97764	Adq97764 Human can
c 36	210	42.0	110000	13 ADS99457_0	AdS99457 Human kin
37	209.8	42.0	2085	13 ADRO7789	Adr07789 Full leng
38	209.8	42.0	6867	5 ABAL4921	Abal4921 Human ner
39	209.8	42.0	7270	5 ABAL4924	Abal4924 Human ner
40	209.8	42.0	110000	11 ACN45090_1	Continuation (2 of
41	209.8	42.0	110000	11 ACN45090_2	Continuation (3 of
42	209.4	41.9	2183	4 AAHL7953	Aahl7953 Human CDN
c 43	209.2	41.9	27204	11 ACN44866	Acn44866 Human gen
c 44	209	41.8	1990	4 ABK42788	Abk42788 Genomic s
45	209	41.8	1990	4 AAK87346	Aak87346 Human imm

ALIGNMENTS

RESULT 1

ADI35082
ID ADI35082 standard; DNA; 12174 BP.

XX AC ADI35082;

XX DT 22-APR-2004 (first entry)

XX DE Human PLA2G1B nucleotide sequence.

XX KW PLA2G1B ; fat deposition; leanness; polymorphism;

XX KW non-insulin dependent diabetes mellitus; NIDDM; hyperinsulinemia;

XX KW hypertension; glucose intolerance; dyslipidemia; hypercoagulability;

XX KW microalbuminuria; human; gene; ds.

XX OS Homo sapiens.

XX XX WO2004002295-A2.

XX XX 08-JAN-2004.

XX PD 08-JAN-2004.

XX PF 27-JUN-2003; 2003WO-US020830.

XX PR 27-JUN-2002; 2002US-0392361P.

XX PA (SEQL-) SEQUENOM INC.

XX PI Adam GIR, Langdown ML;

XX DR WPI; 2004-082843/08.

XX DR P-PSDB; ADI35083.

XX XX

PT Diagnosing a predisposition to fat deposition or leanness, useful for
PT diagnosing a predisposition to e.g. diabetes or hypertension, comprises
PT detecting the presence of a polymorphism in the PLA2G1B nucleic acid from
XX the subject.

PS Claim 1; SEQ ID NO 1; 91pp; English.

CC The invention relates to diagnosing a predisposition to fat deposition or
CC leanness in a subject comprising detecting the presence or absence of a
CC polymorphic variation associated with fat deposition at a polymorphic

CC site in a PLA2G1B nucleotide sequence in a nucleic acid sample from a
CC subject, where the presence of the polymorphic variation indicates a
CC predisposition to fat deposition in the subject. The polymorphic
CC variation is a guanine at position 7328 or thymine at position 9182 of
CC the present sequence. The method is useful for diagnosing a
CC predisposition to fat deposition or leanness in a subject, and
CC consequently for diagnosing a predisposition to non-insulin dependent
CC diabetes mellitus (NIDDM) in a subject and conditions such as
CC hyperinsulinemia, hypertension, glucose intolerance, dyslipidemia,
CC hypercoagulability, or microalbuminuria, which can lead to early
CC prescription of preventive measures. The present sequence represents a
CC human PLA2G1B nucleotide sequence.

XX SQ Sequence 12174 BP; 3217 A; 2992 C; 2738 G; 3215 T; 0 U; 12 Other;
Query Match 100.0%; Score 499.6; DB 12; Length 12174;
Best Local Similarity 100.0%; Pred. No. 3.3e-146;
Matches 500; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 GACCTACCTCGACCTTGTGCGAGTCTTAGCATATGGACCTGGGATGGATTAGCGC 60
Db |||||
Qy 1 GACCTACCTCGACCTTGTGCGAGTCTTAGCATATGGACCTGGGATGGATTAGCGC 60
Db |||||

Qy 61 TCAGTTAATAGTAACCTATAGCCAGGTGCGGTGCTATGCTATTCCAGCACATT 120
Db |||||

Qy 61 TCAGTTAATAGTAACCTATAGCCAGGTGCGGTGCTATGCTATTCCAGCACATT 120
Db |||||

Qy 121 GGGAGACCGAGTTGGTGATCACTTGAGAGCAGGATTTGAGACGAGCTGGCCAAACAT 180
Db |||||

Qy 121 GGGAGACCGAGTTGGTGATCACTTGAGAGCAGGATTTGAGAGCAGCTGGCCAAACAT 180
Db |||||

Qy 181 GGCAGAACACTATCTCTAATAAAAAATACAAAAATTAGCCAGTGTGGTGCACTTGCCTA 240
Db |||||

Qy 181 GGCAGAACACTATCTCTAATAAAAAATACAAAAATTAGCCAGTGTGGTGCACTTGCCTA 240
Db |||||

Qy 241 TAGTCCAGCTACAGGAGGCTGGGCGAGAGATCACTTGAACCTGGGAGGTGGAGT 300
Db |||||

Qy 241 TAGTCCAGCTACAGGAGGCTGGGCGAGAGATCACTTGAACCTGGGAGGTGGAGT 300
Db |||||

Qy 301 TGCAGTGAGCCAGATTTGACCACTGCCTCCAGCTGGAAAAAAGGTAATTAATAC 360
Db |||||

Qy 301 TGCAGTGAGCCAGATTTGACCACTGCCTCCAGCTGGAAAAAAGGTAATTAATAC 360
Db |||||

Qy 361 TTTACTTGAACCATAGCTGCTTCTCTTCTTTTGAGCCACCCCAATCACCCTTAGCA 420
Db |||||

Qy 361 TTTACTTGAACCATAGCTGCTTCTCTTCTTTTGAGCCACCCCAATCACCCTTAGCA 420
Db |||||

Qy 421 TCCTTCAGGCTAAAYCTAGGAGCAGTGCCTGCTCTGCTTGTATGACCCCAAGGA 480
Db |||||

Qy 421 TCCTTCAGGCTAAAYCTAGGAGCAGTGCCTGCTCTGCTTGTATGACCCCAAGGA 480
Db |||||

Qy 481 ACCACATAGAGGGACTGA 500
Db |||||

Qy 481 ACCACATAGAGGGACTGA 500
Db |||||

RESULT 2
ADJ09983
ID ADJ09983 standard; DNA; 12174 BP.
XX
AC ADJ09983;
XX
DT 17-JUN-2004 (first entry)
XX
DE Human phospholipase A2 (PLA2G1B) DNA SeqID 1.
XX human; gene; ds; fat reduction; fat deposition; phospholipase A2;
KW PLA2G1B; chromosome 12q24; single nucleotide polymorphism; SNP;
KW appetite suppressant; lipase inhibitor; exercise regimen; obesity;
KW non-insulin dependent diabetes mellitus; NIDDM; cardiovascular disorder;
KW hypertension; antidiabetic.
XX Homo sapiens.
OS

XX PH Location/Qualifiers
FT variation replace(436,c)
FT /*tag= a
FT /standard_name= "Single nucleotide polymorphisms"
FT variation replace(839,a)
FT /*tag= b
FT /standard_name= "Single nucleotide polymorphisms"
FT variation replace(4050,a)
FT /*tag= c
FT /standard_name= "Single nucleotide polymorphisms"
FT variation replace(4689,t)
FT /*tag= d
FT /standard_name= "Single nucleotide polymorphisms"
FT variation replace(6282,a)
FT /*tag= e
FT /standard_name= "Single nucleotide polymorphisms"
FT variation replace(6358,c)
FT /*tag= f
FT /standard_name= "Single nucleotide polymorphisms"
FT variation replace(6653,t)
FT /*tag= g
FT /standard_name= "Single nucleotide polymorphisms"
FT variation replace(7256,t)
FT /*tag= h
FT /standard_name= "Single nucleotide polymorphisms"
FT variation replace(7300,a)
FT /*tag= i
FT /standard_name= "Single nucleotide polymorphisms"
FT variation replace(7301,c)
FT /*tag= j
FT /standard_name= "Single nucleotide polymorphisms"
FT variation replace(7328,g)
FT /*tag= k
FT /standard_name= "Single nucleotide polymorphisms"
FT variation replace(8062,c)
FT /*tag= l
FT /standard_name= "Single nucleotide polymorphisms"
FT variation replace(9182,t)
FT /*tag= m
FT /standard_name= "Single nucleotide polymorphisms"
FT variation replace(11649,c)
FT /*tag= n
FT /standard_name= "Single nucleotide polymorphisms"
XX WO2004002296-A2.
XX 08-JAN-2004.
XX 27-JUN-2003; 2003WO-US020831.
XX 27-JUN-2002; 2002US-0392362P.
XX (SEQU-) SEQUENOM INC.
XX Adam GIR, Langdown ML, Denissenko MF, Dennis E, Cantor C;
XX Rubin B;
XX WPI; 2004-071944/07.
XX P-PSDB; ADJ09984.
XX Identifying a candidate therapeutic for fat reduction, useful for
XX treating diabetes, by introducing a test molecule to a system comprising
XX PLA2G1B protein or nucleic acid, and determining the presence of
XX interaction between the compounds.
XX Claim 1; SEQ ID NO 1; 116pp; English.
XX This invention relates to a novel candidate therapeutic agent useful for
XX fat reduction and disorders related to fat depositions. Specifically, it
XX refers to polymorphic variations in the phospholipase A2 (PLA2G1B) DNA,
XX which is located on chromosome 12q24 and has been associated with central
XX fat deposition. The present invention describes methods to detect the

CC presence or absence of these single nucleotide polymorphisms of PLA2G1B,
 CC in particular G7328A and T9182G, and subsequently provide treatment that
 CC reduces fat deposition. This treatment may consist of an appetite
 CC suppressant, a lipase inhibitor, a phospholipase inhibitor, an exercise
 CC regimen, a dietary regimen, psychological counselling, psychotherapy or a
 CC psychotherapeutic. Accordingly, PLA2G1B is a target for reducing fat
 CC deposition and it can be used to treat both obesity and non-insulin
 CC dependent diabetes mellitus (NIDDM), as well as cardiovascular disorders
 CC such as hypertension. As such, it exhibits antidiabetic activity. This
 CC polynucleotide sequence is the human PLA2G1B DNA of the invention.
 XX
 SQ Sequence 12174 BP; 3220 A; 2996 C; 2739 G; 3219 T; 0 U; 0 Other;

Query Match 100.0%; Score 499.6; DB 12; Length 12174;
 Best Local Similarity 99.8%; Pred. No. 3.3e-146;
 Matches 499; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 1 GACCTACCTCGACCTTTGGCCAGGTTCTTACATATGGGACCTGGGATGGAGTTAGCGC 60
 DB 1 GACCTACCTCGACCTTTGGCCAGGTTCTTACATATGGGACCTGGGATGGAGTTAGCGC 60
 QY 61 TCAGTTAATAGTAACCTCATTAGCCAGGTGCGTGCATGCTGTTATCCAGCACATT 120
 DB 61 TCAGTTAATAGTAACCTCATTAGCCAGGTGCGTGCATGCTGTTATCCAGCACATT 120
 QY 121 GGGAGACCCAGTTGGGTGGATCACTTGAGAGCAGGAGTTTGAGACAGCCTGGCCAAACAT 180
 DB 121 GGGAGACCCAGTTGGGTGGATCACTTGAGAGCAGGAGTTTGAGACAGCCTGGCCAAACAT 180
 QY 181 GCGAAACACTATCTCTAATAAATAACAAAATTTAGCAGGTGTTGGGCACTTGCCTA 240
 DB 181 GCGAAACACTATCTCTAATAAATAACAAAATTTAGCAGGTGTTGGGCACTTGCCTA 240
 QY 241 TAGTCCAGCTACAGAGGCTGGGCGAGAGATCACTTGAACCTGGGAGTTGAGGT 300
 DB 241 TAGTCCAGCTACAGAGGCTGGGCGAGAGATCACTTGAACCTGGGAGTTGAGGT 300
 QY 301 TGCAGTGAGCCAGGATTCACCACTGCACTCAGGCTGGAAAAAGGTAATTAATAC 360
 DB 301 TGCAGTGAGCCAGGATTCACCACTGCACTCAGGCTGGAAAAAGGTAATTAATAC 360
 QY 361 TTACTTGGAAACCATAGCTGCTTCTCTTTTGGAGCCACCCCAATCACCACATTAGCA 420
 DB 361 TTACTTGGAAACCATAGCTGCTTCTCTTTTGGAGCCACCCCAATCACCACATTAGCA 420
 QY 421 TCCTTCAGGCTTAAATCTAGGAGCAGTGCCTGCTCTGTTATGACCCCAAGGA 480
 DB 421 TCCTTCAGGCTTAAATCTAGGAGCAGTGCCTGCTCTGTTATGACCCCAAGGA 480
 QY 481 ACCACATAAGAGGGACTGA 500
 DB 481 ACCACATAAGAGGGACTGA 500

RESULT 3
 ABN61073
 ID ABN61073 standard; cDNA; 662 BP.
 XX
 AC ABN61073;
 XX
 DT 28-JUN-2002 (first entry)
 XX
 DE Human cancer related polynucleotide SEQ ID NO 1040.
 XX
 KW Human; cytostatic; gene expression; gene mapping; tissue profiling;
 KW gene therapy; cancer; tumour; gene; ss.
 XX
 OS Homo sapiens.
 XX
 FN WO200214500-A2.
 XX
 PD 21-FEB-2002.
 XX

PF 16-AUG-2001; 2001WO-US025840.
 XX
 PR 16-AUG-2000; 2000US-0226326P.
 XX
 PA (CHIR) CHIRON CORP.
 PA (HYSE-) HYSEQ INC.
 XX
 PI Escobedo J, Garcia PD, Sudduth-Klinger J, Reinhard C, Randazzo F;
 PI Lamson G, Scott EM, Zhang G, Kassam A, Pot D, Labat I;
 XX
 DR WPI; 2002-241905/29.
 XX
 PT New nucleic acid for producing a polypeptide, detecting differentially
 PT expressed genes correlated with a cancerous state of a mammalian cell,
 PT and inhibiting tumor growth.
 XX
 PS Claim 1; SEQ ID NO 1040; 883pp + Sequence Listing; English.
 XX
 CC The invention relates to an isolated polynucleotide (ABN27253-ABN33262)
 CC with cytostatic activity. The polynucleotide is used to produce a
 CC polypeptide, to detect differentially expressed genes correlated with a
 CC cancerous state of a mammalian cell and to inhibit tumour growth. The
 CC polynucleotide is used as a probe in mapping and tissue profiling. The
 CC encoded polypeptide and antibodies to the polypeptide can also be used
 CC for therapeutic and diagnostic purposes. The polynucleotide is useful for
 CC gene therapy. Note: The sequence data for this patent did not form part
 CC of the printed specification, but was obtained in electronic format
 CC directly from WIPO at ftp.wipo.int/pub/published_pct_sequences
 XX
 SQ Sequence 662 BP; 202 A; 136 C; 174 G; 150 T; 0 U; 0 Other;

Query Match 43.0%; Score 214.8; DB 6; Length 662;
 Best Local Similarity 83.8%; Pred. No. 5.4e-57;
 Matches 243; Conservative 0; Mismatches 47; Indels 0; Gaps 0;
 QY 63 AGTTAATAGTAACCTCATTAGCCAGGTGCGTGCATGCTGTTATCCAGCACATTGG 122
 DB 184 AGTAAAGAGTTTGGATTGGCCAGGACGGTGGCTCACTCTGTAATCCAGCACATTGG 243
 QY 123 GAGACCCAGTTGGTGGATCACTTGAGAGCAGGAGTTTGAGACAGCCTGGCCAAACATGG 182
 DB 244 GAGCCGAGGCGAGGTGGATCACTTGAGGTGAGGAGTTTGAGACAGGCTGGCCAAACATGG 303
 QY 183 CAAACACTATCTCTAATAAATAACAAAATTTAGCAGGTGTTGGGCACTTGCCTATA 242
 DB 304 TGAACCTGCTCTCTAATAAATAACAAAATTTAGCAGGTTGTTGGCAGCCCTGTA 363
 QY 243 GTCCAGCTACACAGGAGGCTGGGGCAGAGAATCACTTGAACCTGGGAGGTGAGGTTG 302
 DB 364 GCCCAGCTACTCTGGCGGCTGAGGCGAGGAGATCGCTTGAACCTGGGAGGTGAGGCTG 423
 QY 303 CAGTGAGCCAGGATTCACCACTGCACTCAGCCTGGAAAAAGGTTAA 352
 DB 424 CAGTGAGCTGAGATCGCACCACCTGCATTCAGCCTGGGCAACAAAGTGA 473

RESULT 4
 ACL58180
 ID ACL58180 standard; cDNA; 662 BP.
 XX
 AC ACL58180;
 XX
 DT 24-MAR-2005 (first entry)
 XX
 DE Human colon cancer differentially expressed polynucleotide, SEQ ID:4315.
 KW Differential expression; diagnosis; therapy; drug screening; cancer;
 KW neoplasm; colon tumor; breast tumor; pancreas tumor; cytostatic; vaccine;
 KW ss.
 XX
 OS Homo sapiens.
 XX
 FN WO200500087-A2.
 XX

XX 06-JAN-2005.
 PD 13-MAY-2004; 2004WO-US015421.
 XX 03-JUN-2003; 2003US-0475872P.
 XX (CHIR) CHIRON CORP.
 PA Randazzo F, Moler E, Escobedo J, Garcia PD;
 PI WPI; 2005-075421/08.
 XX New isolated polynucleotides, which are differentially expressed in colon
 PT cancer cell, useful for treating cancer, e.g. colon cancer, breast
 PT cancer, or pancreatic cancer.
 XX
 PS Claim 1; SEQ ID NO 4315; 97pp; English.
 XX
 CC The invention relates to 9672 polynucleotides (ACLS3866-ACL63537) which
 CC are differentially expressed in colon cancer cells. The invention also
 CC relates to vectors and host cells comprising a differentially expressed
 CC polynucleotide of the invention; a method for detecting a cancerous cell
 CC by detection of a gene product of the polynucleotides; a method for
 CC inhibiting a cancerous phenotype of a cell by inhibiting a gene product
 CC of the polynucleotides; a method of treating an individual with cancer by
 CC administration of a modulator of a gene product of the polynucleotides;
 CC and an isolated antibody that specifically binds to a polypeptide encoded
 CC by one of the 9672 polynucleotides. The polynucleotides, polypeptides,
 CC antibodies, and methods are useful for the detection of cancerous cells;
 CC for the diagnosis, prognosis and management of cancer; for the
 CC identification of agents that modulate the phenotype of cancerous cells;
 CC for the identification of therapeutic targets for cancer chemotherapy;
 CC and for the treatment of cancer, especially colon cancer and metastasized
 CC colon cancer, but also breast or pancreatic cancer. The polynucleotides
 CC are also useful as a source of probes or primers for use in diagnostic
 CC methods. The differentially expressed polynucleotides or their encoded
 CC proteins can additionally be used as vaccines to modulate primary immune
 CC responses for the prevention or treatment of cancer. The present invention
 CC represents a specifically claimed polynucleotide which is differentially
 CC expressed in colon cancer. Note: The sequence data for this patent did
 CC not form part of the printed specification, but was obtained in
 CC electronic format directly from WIPO at
 CC ftp.wipo.int/pub/published_pct_sequences
 XX
 SQ Sequence 662 BP; 202 A; 136 C; 174 G; 150 T; 0 U; 0 Other;
 Query Match 43.0%; Score 214.8; DB 14; Length 662;
 Best Local Similarity 83.8%; Pred. No. 5.4e-57;
 Matches 243; Conservative 0; Mismatches 47; Indels 0; Gaps 0;
 QY 63 AGTTAATAGTAACCTATTAGCCAGGTGCGGTGCTCATGTCTGTATTCACGACTTTGG 122
 DB 184 AGTAAAGAGTTTGGATTGGCCAGGACGCGTGGCTCACTCTGTATTCACGACTTTGG 243
 QY 123 GAGACCGAGTTGGTGTGATCACTTGAGCAGAGAGTTTGAGACCGCTGCGCAATCG 182
 DB 244 GAGGCCGAGCGAGGTGGATCACTTGAGGTCAGAGTTTCAGACCGCTGCGCAATCG 303
 QY 183 CAAACACTATCTCTAATAAATAACAAAATAGCCAGGTGTGGTGGCACTTGCCTATA 242
 DB 304 TGAACCCCTGTCTCTAATAAATAACAAAATAGCCAGGTGTGGTGGCACTTGCCTATA 363
 QY 243 GTCCAGCTACACAGGAGGCTGGGCGAGAGAAATCACTTGAACCTGGGAGGTGGAGTTG 302
 DB 364 GCCCAGCTACTCTGGCGGTGAGGCGAGGAGATCGTTGAACCTGGGAGGTGGAGCTG 423
 QY 303 CAGTGACCGCAAGATTGCACCACTGCCTCAGGCTCGCAAAAAGGTTAA 352
 DB 424 CAGTGAGCTGAGATCGCACCACCTGCATTCAGGCTTGGGCAACAAAAGTGA 473

ADZ59751
 ID ADZ59751 standard; DNA; 35804 BP.
 XX
 AC ADZ59751;
 XX
 DT 14-JUL-2005 (first entry)
 XX
 DE Human CB1 cannabinoid receptor interacting protein 1b gene SEQ ID NO:3.
 XX
 KW CB1 cannabinoid receptor interacting protein 1; antiemetic; anabolic;
 KW analgesic; inotropic; ophthalmological; gastrointestinal-gen.; nootropic;
 KW neuroprotective; analgesia; multiple sclerosis; nausea; drug-induced;
 KW cancer; chemotherapy; retinoblastoma; chromosome-2; gene; ds.
 XX
 OS Homo sapiens.
 XX
 FN US2005095674-A1.
 XX
 PD 05-MAY-2005.
 XX
 XX 23-JUL-2004; 2004US-00898406.
 XX
 XX 23-JUL-2003; 2003US-0489542P.
 PR 27-FEB-2004; 2004US-0548697P.
 XX
 PA (LEWI/) LEWIS D L.
 PA (BHAR/) BHARTUR S.
 PA (WALL/) WALLIS K.
 PA (NIEH/) NIEHAUS J.
 XX
 PI Lewis DL, Bhartur S, Wallis K, Niehaus J;
 XX
 DR WPI; 2005-332074/34.
 XX
 PT Novel isolated cannabinoid receptor nucleic acid, useful for modulating
 PT activity of CB1 cannabinoid receptor such as appetite stimulation,
 PT analgesia, euphoria, decreased tremor or spasticity associated with
 PT multiple sclerosis.
 XX
 PS Claim 1; SEQ ID NO 3; 77pp; English.
 XX
 CC The invention relates to an isolated CB1 cannabinoid receptor interacting
 CC protein 1 (CRIP1) nucleic acid (I). Also described: (1) a vector
 CC comprising (I); (2) a composition (II) comprising (I) and a carrier; (3)
 CC modulating the activity of a CB1 cannabinoid receptor comprising
 CC administering (II) to the CB1 cannabinoid receptor; (4) an antibody that
 CC specifically binds to a CRIP1 polypeptide; and (5) screening for a
 CC modulator of CRIP1 activity on a CB1 cannabinoid receptor. (I) and (II)
 CC are useful for modulating the activity of a CB1 cannabinoid receptor,
 CC which involves administering (I) or (II) to the CB1 cannabinoid receptor
 CC in a cell, which is a human cell of a patient. The modulation of CB1
 CC cannabinoid receptor in the patient results in one or more effects chosen
 CC from appetite stimulation, analgesia, euphoria, decreased tremor or
 CC spasticity associated with multiple sclerosis, attenuation of nausea and
 CC vomiting in cancer, chemotherapy, reduction in intraocular pressure,
 CC decreased intestinal motility, and attenuation of aversive memories,
 CC appetite suppression, decreased sedation, decreased alterations in
 CC cognition and memory, and decreased mood alterations. (II) is useful for
 CC modulating the tonic activity of the CB1 cannabinoid receptor, which
 CC involves administering (II) to the CB1 cannabinoid receptor. Human CRIP1
 CC and CRIP1b map to chromosome 2, and are retinoblastoma genes. The present
 CC sequence represents the human CRIP1b genomic DNA sequence.
 XX
 SQ Sequence 35804 BP; 11398 A; 7034 C; 7287 G; 10085 T; 0 U; 0 Other;
 Query Match 43.0%; Score 214.8; DB 14; Length 35804;
 Best Local Similarity 83.8%; Pred. No. 2.9e-56;
 Matches 243; Conservative 0; Mismatches 47; Indels 0; Gaps 0;
 QY 63 AGTTAATAGTAACCTATTAGCCAGGTGCGGTGCTCATGTCTGTATTCACGACTTTGG 122
 DB 25857 AGTAAAGAGTTTGGATTGGCCAGGACGCGTGGCTCACTCTGTATTCACGACTTTGG 25916

	Beet	Local Similarity	78.3%;	Pred. No. 1.4e-56;	Mismatches	0;	Mismatches	71;	Indels	0;	Gaps	0;
Qy	26	TTC	TAGCATATGGGACCTGGGATGGAGTTAGCGCTCAGTTAAATAGTAACACTCATTTAGCCCA	85								
Db	255	TTTTT	TATGATGTGAGTCCCAAGATAGTTTACAATGATAAAGAGGAAGAAGCTGGCGTG	314								
Qy	86	GGT	CGGTTGGCTCATGTCGTATTCCAGCACACTTTGGGAGNCCAGATTTGGTGGATCACT	145								
Db	315	GGCC	CAGTGGCTCAAGCTGTAAATCCACGACACTTTGGAGGCTGAGCGGGTGGATCACC	374								
Qy	146	TGAG	CAGCAGGATTTGAGACGAGCTGGCCAACATGSCAAACACTATCTCTATAAAAA	205								
Db	375	TGAG	CTCAGGAGTTTGAGACGAGCTTGGCAAACGTCGCCAAACCCCGTCTCTACTAAAAA	434								
Qy	206	TACAAA	AATTAGCCAGGTGTGGTGGCACTTGCCTTATAGTCCCAGCTACACAGGAGGTGG	265								
Db	435	TACAAA	AATTTAGCCGGCGGTGTGGCGGGCGCCTGTAAATCCAGCTACTCAGGAGGTGA	494								
Qy	266	GGCAGA	AGAATCACTTGAACCTGGGAGGTGGAGGTTCGAGTGAAGCCAAAGATTGACCACT	325								
Db	495	GGCAGG	AGAAATGCTTGAACCCGGAGGCGGAGGTTCGAGTGAAGCCAAAGATTGCGCCATT	554								
Qy	326	GCATC	CAGCGCTGGAIAAAAAAGGGTAA	352								
Db	555	GCATC	CAGCGCTGGCAACCAAGATGA	581								

RESULT 8	
ACL57703	
ID	ACL57703 standard; cDNA; 587 BP.
XX	
AC	ACL57703;
AC	
XX	
DT	24-MAR-2005 (first entry)
XX	
DE	Human colon cancer differentially expressed polynucleotide, SEQ ID:3838.
DE	
XX	
KW	Differential expression; diagnosis; therapy; drug screening; cancer;
KW	neoplasia; colon tumor; breast tumor; pancreas tumor; cytostatic; vaccine;
XX	
XX	ss.
OS	Homo sapiens.
OS	
XX	
PN	WO2005000087-A2.
XX	
PD	06-JAN-2005.
XX	
PF	13-MAY-2004; 2004WO-US015421.
XX	
XX	
PR	03-JUN-2003; 2003US-0475872P.
XX	
PA	(CHIR) CHIRON CORP.
XX	
PI	Randazzo F, Moler E, Escobedo J, Garcia PD;
XX	
XX	WPI; 2005-075421/08.
XX	
PT	New isolated polynucleotides, which are differentially expressed in colon
PT	cancer cell, useful for treating cancer, e.g. colon cancer, breast
PT	cancer, or pancreatic cancer.
XX	
XX	Claim 1; SEQ ID NO 3838; 97pp; English.

by one of the 9672 polynucleotides. The polynucleotides, polypeptides, antibodies, and methods are useful for the detection of cancerous cells; for the diagnosis, prognosis and management of cancer; for the identification of agents that modulate the phenotype of cancerous cells; for the identification of therapeutic targets for cancer chemotherapy; and for the treatment of cancer, especially colon cancer and metastasized colon cancer, but also breast or pancreatic cancer. The polynucleotides are also useful as a source of probes or primers for use in diagnostic methods. The differentially expressed polynucleotides or their encoded proteins can additionally be used as vaccines to modulate primary immune responses for the prevention or treatment of cancer. The present sequence represents a specifically claimed polynucleotide which is differentially expressed in colon cancer. Note: The sequence data for this patent did not form part of the printed specification, but was obtained in electronic format directly from WIPO at [ftp.wipo.int/pub/published](http://wipo.int/pub/published) pct sequences

Seq	Sequence	587 BP	149 A	131 C	163 G	144 T	0 U	0 Other
Query	Match	42.7%	Score	213.4	DB	14	Length	587
Best Local Similarity	78.3%	Pred.	No.	1.4e-56				
Matches	256	Conservative	0	Mismatches	71	Indels	0	Gaps
Qy	26	TTCTTAGCATATGGACCTGGATGGAGTTAGCGCTCAGTTAAATAGTAACCTCATTTAGCCA	85					
Db	255	TTTTTATGATGTGAGTCCCAAGATAGTTTACAAATGATAAAGAGGAAGCTGGCGTG	314					
Qy	86	GGTCGGCTGGCTCATGTCGTATTCCAGACACTTTGGGAGACCCAGTTGGGTGGATCACT	145					
Db	315	GGCGCAGTGGCTCAGCGCTGTATCCAGACACTTTGGAGGCTGAGCGGGTGGATCAC	374					
Qy	146	TGAGAGCAGGAGTTTGAGACAGCGCTGGCCAACATGGCAAACACTATCTCTAATAAAAA	205					
Db	375	TGAGGTCAGGAGTTTGAGACAGCGCTTGGCCAAACGTGGCGAAACCCCGTCTCTACTAAAA	434					
Qy	206	TACAAAAATTAGCCAGGTGTGGTGGCACTTGCCTATAGTCCAGGTACACAGGAGGCTGG	265					
Db	435	TACAAAAATTAGCCGGCGGTGGTGGCGGGCGCCTGTAAATCCAGCTACTCAGGAGGCTGA	494					
Qy	266	GGCAGAGAAGATCACTTGAACCTGGGAGGTGGAGGTTCGAGTGGACCAAGATTGGACCACT	325					
Db	495	GGCAGGAGAAATTCGTTGAAACCCGGAGCGGAGGTTGCAGTGAAGCCAGATTGGCCATT	554					
Qy	326	GCACCTCAGCGCTGAAAAAAGGGTAA	352					
Db	555	GCACCTCAGCGCTGGCAACCAAGAGTGA	581					

RESULT 9	
AAK79514	
ID	AAK79514 standard; DNA; 9469 BP.
XX	
XX	AAK79514;
XX	
XX	07-NOV-2001 (first entry)
DT	
XX	
DE	Human immune/haematopoietic antigen genomic sequence SEQ ID NO:34326.
DE	
XX	
KW	Human; immune; haematopoietic; immune/haematopoietic antigen; cancer;
KW	cytostatic; gene therapy; vaccine; metastasis; ds.
XX	
OS	Homo sapiens.
XX	
XX	WO200157182-A2.
XX	
PD	09-AUG-2001.
PF	
XX	
XX	17-JAN-2001; 2001WO-US001354.
XX	
PR	31-JAN-2000; 2000US-0179065P.
PR	04-FEB-2000; 2000US-0180628P.
PR	24-FEB-2000; 2000US-0184664P.
PR	02-MAR-2000; 2000US-0186350P.

PR 16-MAR-2000; 2000US-0189874P.
PR 17-MAR-2000; 2000US-0190076P.
PR 18-APR-2000; 2000US-0198123P.
PR 19-MAY-2000; 2000US-0205515P.
PR 07-JUN-2000; 2000US-0209467P.
PR 28-JUN-2000; 2000US-0214886P.
PR 30-JUN-2000; 2000US-0215135P.
PR 07-JUL-2000; 2000US-0216647P.
PR 07-JUL-2000; 2000US-0216880P.
PR 11-JUL-2000; 2000US-0217487P.
PR 11-JUL-2000; 2000US-0217496P.
PR 14-JUL-2000; 2000US-0218290P.
PR 26-JUL-2000; 2000US-0220963P.
PR 26-JUL-2000; 2000US-0220964P.
PR 14-AUG-2000; 2000US-0224518P.
PR 14-AUG-2000; 2000US-0224519P.
PR 14-AUG-2000; 2000US-0225211P.
PR 14-AUG-2000; 2000US-0225214P.
PR 14-AUG-2000; 2000US-0225266P.
PR 14-AUG-2000; 2000US-0225267P.
PR 14-AUG-2000; 2000US-0225268P.
PR 14-AUG-2000; 2000US-0225270P.
PR 14-AUG-2000; 2000US-0225447P.
PR 14-AUG-2000; 2000US-0225757P.
PR 14-AUG-2000; 2000US-0225758P.
PR 14-AUG-2000; 2000US-0225759P.
PR 18-AUG-2000; 2000US-0226279P.
PR 22-AUG-2000; 2000US-0226681P.
PR 22-AUG-2000; 2000US-0226868P.
PR 22-AUG-2000; 2000US-0227182P.
PR 23-AUG-2000; 2000US-0227009P.
PR 30-AUG-2000; 2000US-0228924P.
PR 01-SEP-2000; 2000US-0229287P.
PR 01-SEP-2000; 2000US-0229343P.
PR 01-SEP-2000; 2000US-0229344P.
PR 01-SEP-2000; 2000US-0229345P.
PR 05-SEP-2000; 2000US-0229509P.
PR 05-SEP-2000; 2000US-0229513P.
PR 06-SEP-2000; 2000US-0230437P.
PR 06-SEP-2000; 2000US-0230438P.
PR 08-SEP-2000; 2000US-0231242P.
PR 08-SEP-2000; 2000US-0231243P.
PR 08-SEP-2000; 2000US-0231244P.
PR 08-SEP-2000; 2000US-0231413P.
PR 08-SEP-2000; 2000US-0231414P.
PR 08-SEP-2000; 2000US-0232080P.
PR 08-SEP-2000; 2000US-0232081P.
PR 12-SEP-2000; 2000US-0231968P.
PR 14-SEP-2000; 2000US-0232397P.
PR 14-SEP-2000; 2000US-0232398P.
PR 14-SEP-2000; 2000US-0232399P.
PR 14-SEP-2000; 2000US-0232400P.
PR 14-SEP-2000; 2000US-0232401P.
PR 14-SEP-2000; 2000US-0233663P.
PR 14-SEP-2000; 2000US-0233664P.
PR 14-SEP-2000; 2000US-0233665P.
PR 21-SEP-2000; 2000US-0234223P.
PR 21-SEP-2000; 2000US-0234274P.
PR 25-SEP-2000; 2000US-0234997P.
PR 25-SEP-2000; 2000US-0234998P.
PR 26-SEP-2000; 2000US-0235484P.
PR 27-SEP-2000; 2000US-0235834P.
PR 27-SEP-2000; 2000US-0235836P.
PR 29-SEP-2000; 2000US-0236327P.
PR 29-SEP-2000; 2000US-0236367P.
PR 29-SEP-2000; 2000US-0236368P.
PR 29-SEP-2000; 2000US-0236369P.
PR 29-SEP-2000; 2000US-0236370P.
PR 02-OCT-2000; 2000US-0236802P.
PR 02-OCT-2000; 2000US-0237037P.
PR 02-OCT-2000; 2000US-0237038P.
PR 02-OCT-2000; 2000US-0237039P.
PR 02-OCT-2000; 2000US-0237040P.

PR 13-OCT-2000; 2000US-0239935P.
PR 13-OCT-2000; 2000US-0239937P.
PR 20-OCT-2000; 2000US-0240960P.
PR 20-OCT-2000; 2000US-0241221P.
PR 20-OCT-2000; 2000US-0241785P.
PR 20-OCT-2000; 2000US-0241786P.
PR 20-OCT-2000; 2000US-0241787P.
PR 20-OCT-2000; 2000US-0241808P.
PR 20-OCT-2000; 2000US-0241809P.
PR 20-OCT-2000; 2000US-0241826P.
PR 01-NOV-2000; 2000US-0244617P.
PR 08-NOV-2000; 2000US-0246474P.
PR 08-NOV-2000; 2000US-0246475P.
PR 08-NOV-2000; 2000US-0246476P.
PR 08-NOV-2000; 2000US-0246477P.
PR 08-NOV-2000; 2000US-0246478P.
PR 08-NOV-2000; 2000US-0246523P.
PR 08-NOV-2000; 2000US-0246524P.
PR 08-NOV-2000; 2000US-0246525P.
PR 08-NOV-2000; 2000US-0246526P.
PR 08-NOV-2000; 2000US-0246527P.
PR 08-NOV-2000; 2000US-0246528P.
PR 08-NOV-2000; 2000US-0246532P.
PR 08-NOV-2000; 2000US-0246610P.
PR 08-NOV-2000; 2000US-0246611P.
PR 08-NOV-2000; 2000US-0246613P.
PR 17-NOV-2000; 2000US-0249207P.
PR 17-NOV-2000; 2000US-0249208P.
PR 17-NOV-2000; 2000US-0249209P.
PR 17-NOV-2000; 2000US-0249210P.
PR 17-NOV-2000; 2000US-0249211P.
PR 17-NOV-2000; 2000US-0249212P.
PR 17-NOV-2000; 2000US-0249213P.
PR 17-NOV-2000; 2000US-0249214P.
PR 17-NOV-2000; 2000US-0249215P.
PR 17-NOV-2000; 2000US-0249216P.
PR 17-NOV-2000; 2000US-0249217P.
PR 17-NOV-2000; 2000US-0249218P.
PR 17-NOV-2000; 2000US-0249244P.
PR 17-NOV-2000; 2000US-0249245P.
PR 17-NOV-2000; 2000US-0249264P.
PR 17-NOV-2000; 2000US-0249265P.
PR 17-NOV-2000; 2000US-0249297P.
PR 17-NOV-2000; 2000US-0249299P.
PR 17-NOV-2000; 2000US-0249300P.
PR 01-DEC-2000; 2000US-0250160P.
PR 01-DEC-2000; 2000US-0250391P.
PR 05-DEC-2000; 2000US-0251030P.
PR 05-DEC-2000; 2000US-0251988P.
PR 05-DEC-2000; 2000US-0256719P.
PR 06-DEC-2000; 2000US-0251479P.
PR 08-DEC-2000; 2000US-0251856P.
PR 08-DEC-2000; 2000US-0251868P.
PR 08-DEC-2000; 2000US-0251869P.
PR 08-DEC-2000; 2000US-0251989P.
PR 08-DEC-2000; 2000US-0251990P.
PR 11-DEC-2000; 2000US-0254097P.
PR 05-JAN-2001; 2001US-0259678P.

(HUMA-) HUMAN GENOME SCI INC.
Rosen CA, Barash SC, Ruben SM;
WPI; 2001-483426/52.

Nucleic acids encoding human immune/hematopoietic antigen polypeptides,
useful for preventing, diagnosing and/or treating cancers and metastasis.
Disclosure; SEQ ID NO 34326; 3071pp + Sequence Listing; English.
AAK54951 to AAK64702 encode the human immune/haematopoietic antigen (I)
amino acid sequences given in AAM82170 to AAM91921. (I) have cytostatic

CC activity, and can be used in gene therapy and vaccine production. (I)
 CC treatment of diseases associated with inappropriate (I) expression. For
 CC example, they may be used to treat disorders associated with decreased
 CC expression by rectifying mutations or deletions in a patient's genome
 CC that affect the activity of (I) by expressing inactive proteins or to
 CC supplement the patients own production of (I). Additionally, (I)
 CC polynucleotides may be used to produce the secreted (I), by inserting the
 CC nucleic acids into a host cell and culturing the cell to express the
 CC protein. (I) proteins and polynucleotides may be used to prevent,
 CC diagnose and treat immune/haematopoietic-related diseases, especially
 CC cancers and cancer metastases of haematopoietic-derived cells. AAK64703
 CC to AAK87694 represent human immune/haematopoietic antigen genomic
 CC sequences from the present invention. AAK54942 to AAK54950 and AAK82169
 CC represent sequences used in the exemplification of the present invention
 XX
 SQ Sequence 9469 BP; 2940 A; 2203 C; 1941 G; 2385 T; 0 U; 0 Other;

Query Match 42.7%; Score 213.4; DB 4; Length 9469;
 Best Local Similarity 84.0%; Pred. No. 4.6e-56;
 Matches 241; Conservative 0; Mismatches 46; Indels 0; Gaps 0;
 QY 66 TAATAGTAACCTATTAGCCAGTGGGCTCATGTCTGTATTCACGACCTTTGGGAG 125
 Db 1143 TGACGGAATAGTTCGGCCAGCGAGGTGGCTCATGCTGTAAATCCAGACCTTTGGGAG 1202
 QY 126 ACCGAGTTGGGTGGATCACTTTGAGAGCAGGAGTTTGAGACCAGCTGGCCAACTGGCAA 185
 Db 1203 GCCAAGTGGGAGGATCACTTGAGTTCAGAGTTTCAGACCAGCTGGCCAACTGGTGA 1262
 QY 186 AACACTATCTCTATATAAAATACAAAATTAGCCAGGTGGTGGCACTTGCCTATAGTC 245
 Db 1263 AACCCCGTCTCTACTATAAAATACAAAATTAGCAGGGCATGGTGGCAGCGCCTGTAATC 1322
 QY 246 CCAGCTACACAGGAGGCTGGGGCAGAGAATCACTTGAACCTGGGAGGTGGAGTTGCAG 305
 Db 1323 CCAGTACTCAGGAGGCTGGAGCAGAGAATCACTTGAACCCGGAGGCGGAGTTGCAG 1382
 QY 306 TGAGCCAGAGTTGCACCACTGCACCTCCAGCTGGAAAAGGGTAA 352
 Db 1383 TGAGCCAGAGTGGCACCCTGCACCTCCAGCTTCAGACCTAGACACAGAGCNA 1429

RESULT 10
 ABT11173
 ID ABT11173 standard; DNA; 168174 BP.
 XX
 AC ABT11173;
 XX
 DT 05-DEC-2002 (first entry)
 XX
 DE Human 5-lipoxygenase gene related DNA sequence SEQ ID No 63.
 XX
 KW Human; polymorphic region; 5-lipoxygenase; 5-LO gene; asthma; bronchitis;
 KW sinusitis; ulcerative colitis; nephritis; amyloidosis; sarcoidosis;
 KW rheumatoid arthritis; scleroderma; lupus; non-allergic rhinitis;
 KW polymyositis; Reiter's syndrome; psoriasis; pelvic inflammatory disease;
 KW atopic; contact dermatitis; forensic medicine; paternity testing; enzyme;
 KW ds.

OS Homo sapiens.
 XX
 XX WO200262825-A2.
 XX
 XX 15-AUG-2002.
 PD
 XX 07-FEB-2002; 2002WO-US003546.
 PF
 XX 08-FEB-2001; 2001US-0267515P.
 PR
 XX 21-AUG-2001; 2001US-0314248P.
 XX
 XX (MILL-) MILLENNIUM PHARM INC.
 PA
 XX

PI Barnes G, Meyer J;
 XX
 DR WPI; 2002-627522/67.
 XX

XX New isolated nucleic acid molecule with an allelic variant of a
 PT polymorphic region of an 5-LO gene, useful for diagnosing and/or
 PT prognosticating disorders associated with an aberrant inflammatory
 PT response such as asthma.

PS Disclosure; Fig 4; 290pp; English.

XX The invention relates to an isolated human nucleic acid molecule
 CC comprising an allelic variant of a polymorphic region of a 5-lipoxygenase
 CC (5-LO) gene, where the allelic variant comprises one or more nucleotide
 CC selected from any of 3, 20 or 21 base pair sequences, given in the
 CC specification, or their complement. The compositions and methods of the
 CC present invention are useful for diagnosing and/or prognosing disorders
 CC associated with an aberrant inflammatory response such as asthma,
 CC bronchitis, sinusitis, ulcerative colitis, nephritis, amyloidosis,
 CC rheumatoid arthritis, sarcoidosis, scleroderma, lupus, non-allergic
 CC rhinitis, polymyositis, Reiter's syndrome, psoriasis, pelvic inflammatory
 CC disease, atopic and contact dermatitis. The nucleic acid molecules can
 CC also be useful for identifying an individual amongst other individuals
 CC from the same species for use in forensic medicine and paternity testing.
 CC This polynucleotide sequence represents DNA relating to the human 5-
 CC lipoxygenase (5-LO) gene of the invention

XX Sequence 168174 BP; 46808 A; 36442 C; 36942 G; 46474 T; 0 U; 1508 Other;

Query Match 42.7%; Score 213.4; DB 6; Length 168174;
 Best Local Similarity 78.3%; Pred. No. 1.15e-55;
 Matches 256; Conservative 0; Mismatches 71; Indels 0; Gaps 0;

QY 26 TTCTTAGCATATGGGACCTGGGATGGAGTTAGCGCTCAGTTAATAGTAATCATTAGCCA 85
 Db 26811 TTTTATGATGTAGTCCCAAGATAGTTTTACAAATGATAAAGAGGAAGCTGGCTG 26870
 QY 86 GGTGCGGTGCTCATGTCTGTTATCCAGCAGCTTTGGGAGACCGAGTTGGTGGATCCT 145
 Db 26871 GGGCGAGTGGCTCAGCGCTGTATCCAGACACTTTGGAGGCTGAGCGGGTGGATCACC 26930
 QY 146 TGAGAGCAGGAGTTTGAGACCAGCTGGCCAAACATGGCAAAACACTATCTTAATAAAAA 205
 Db 26931 TGAGGTTCAGGAGTTTGAGACCAGCTTTGGCCAAACATGGCCGAAACCCCGTCTCTACTAAAAA 26990
 QY 206 TACAAAAATTAGCCAGGTGTGGTGGCACTTGCCTATATAGTCCAGCTACACAGAGGCTGG 265
 Db 26991 TACAAAAATTAGCCGGGCGTGGTGGCGGCGCTGTAAATCCAGCTACTCAGGAGGCTGA 27050
 QY 266 GGCAGAGAATCACTTGAACCTGGGAGGTGGAGTTGCAGTGAGCCAAAGATTCACCACT 325
 Db 27051 GGCAGAGAAATTGCTTGAACCCGGGAGGCGGAGTTGCGATGAGCCCAAGATTGCGCAATT 27110
 QY 326 GCATCTCAGCTGGAAAAAAGGGTAA 352

Db 27111 GCATCTCAGCTGGGCAACAAGAGTGA 27137

RESULT 11
 ABT11114
 ID ABT11114 standard; DNA; 168273 BP.

XX AC ABT11114;

XX DT 05-DEC-2002 (first entry)

XX DE Human 5-lipoxygenase gene related DNA sequence SEQ ID No 2.

XX Human; polymorphic region; 5-lipoxygenase; 5-LO gene; asthma; bronchitis;
 KW sinusitis; ulcerative colitis; nephritis; amyloidosis; sarcoidosis;
 KW rheumatoid arthritis; scleroderma; lupus; non-allergic rhinitis;
 KW polymyositis; Reiter's syndrome; psoriasis; pelvic inflammatory disease;
 KW atopic; contact dermatitis; forensic medicine; paternity testing; enzyme;

ds.
XX Homo sapiens.
XX WO200262825-A2.
XX 15-AUG-2002.
XX 07-FEB-2002; 2002WO-US003546.
XX 08-FEB-2001; 2001US-0267515P.
XX 21-AUG-2001; 2001US-0314248P.
XX (MILL-) MILLENNIUM PHARM INC.
XX Barnes G, Meyer J;
XX WPI; 2002-627522/67.
XX New isolated nucleic acid molecule with an allelic variant of a
XX polymorphic region of an 5-LO gene, useful for diagnosing and/or
XX prognosticating disorders associated with an aberrant inflammatory
XX response such as asthma.
XX Disclosure; Fig 2; 290pp; English.
XX The invention relates to an isolated human nucleic acid molecule
XX comprising an allelic variant of a polymorphic region of a 5-lipoxygenase
XX (5-LO) gene, where the allelic variant comprises one or more nucleotide
XX selected from any of 3, 20 or 21 base pair sequences, given in the
XX specification, or their complement. The compositions and methods of the
XX present invention are useful for diagnosing and/or prognosing disorders
XX associated with an aberrant inflammatory response such as asthma,
XX bronchitis, sinusitis, ulcerative colitis, nephritis, amyloidosis,
XX rheumatoid arthritis, sarcoidosis, scleroderma, lupus, non-allergic
XX rhinitis, polymyositis, Reiter's syndrome, psoriasis, pelvic inflammatory
XX disease, atopic and contact dermatitis. The nucleic acid molecules can
XX also be useful for identifying an individual amongst other individuals
XX from the same species for use in forensic medicine and paternity testing.
XX This polynucleotide sequence represents DNA relating to the human 5-
XX lipoxygenase (5-LO) gene of the invention
XX
XX Sequence 168273 BP; 46834 A; 36467 C; 36966 G; 46498 T; 0 U; 1508 Other;
XX
XX Query Match 42.7%; Score 213.4; DB 6; Length 168273;
XX Best Local Similarity 78.3%; Pred. No. 1.5e-55;
XX Matches 256; Conservative 0; Mismatches 71; Indels 0; Gaps 0;
XX
XX 26 TTCTTAGCATATGGGACCTGGGATGGAGTTAGCGCTCAGTTAATAGTAATCTTAGCCA 85
XX 26861 TTTTATGATGTGAGTCCCAAGATAGTTTACAAATGATAAAGAGGAAGCTGGCTG 26920
XX
XX 86 GGTGGCGTGGCTCATCTGTATCCAGCAGCTTTGGGAGACCGAGTTGGTGGATCACT 145
XX 26921 GGGCAGTGGCTCACCCCTGTATCCAGACATTTGGGAGGCTGAGGCGGTGATCACC 26980
XX
XX 146 TGAGAGCAGGAGTTTGAGACACGCTGGCCACATGGCAAAACACTATCTTAATAAAA 205
XX 26981 TGAGGTGAGGAGTTTGAGACACGCTTTGGCCAGCTGGCGAAACCCCGTCTCTACTAAAA 27040
XX
XX 206 TACAAAAATTAGCCAGTGTGGGCACTTGCTATAGTCCAGCTACACAGGAGCTGG 265
XX 27041 TACAAAAATTAGCCGCGCTGGTGGCGGGCGCTGTAAATCCAGCTACTCAGGAGCTGA 27100
XX
XX 266 GCAGAGAGATCACTTGAACCTGGGAGGTGGAGTTGCGAGTGAGCCAAAGATTGCACCACT 325
XX 27101 GGCAGAGAGATTCCTTGAACCCGGAGGCGGAGGTTGCGAGTGAGCCCAAGATTGCGCAT 27160
XX
XX 326 GCACCTCCAGCTGGAAAAAAGGGTAA 352
XX 27161 GCACCTCCAGCTGGGCAACAAGATGA 27187

RESULT 12

ADQ63479/c
ID ADQ63479 standard; cDNA; 3660 BP.
XX
XX AC ADQ63479;
XX DT 07-OCT-2004 (first entry)
XX DE Novel human cDNA sequence #640.
XX
XX ss; gene; osteopathic; neuroprotective; nootropic; antiparkinsonian;
XX cytosolic; gene therapy; diagnostic marker; morbid state; osteoporosis;
XX neurological disease; Alzheimer's disease; Parkinson's disease; dementia;
XX cancer.
XX
XX Homo sapiens.
XX EPI440981-A2.
XX 28-JUL-2004.
XX 21-JAN-2004; 2004EP-00001196.
XX 21-JAN-2003; 2003JP-00102206.
XX 09-MAY-2003; 2003JP-00131392.
XX (REAS-) RES ASSOC BIOTECHNOLOGY.
XX Isogai T, Sugiyama T, Otsuki T, Wakamatsu A, Sato H, Ishii S;
XX Yamamoto J, Isono Y, Nagai K, Irie R;
XX WPI; 2004-535376/52.
XX P-FSDB; ADQ65667.
XX Novel 2495 cDNA, useful for treating osteoporosis, neurological diseases,
XX Alzheimer's diseases, Parkinson's diseases, dementia and various cancers.
XX Claim 1; SEQ ID NO 640; 2449pp; English.
XX The invention relates to 2495 novel polynucleotides (I) and their encoded
XX polypeptides, sequences hybridizing to these nucleotides, sequences
XX encoding partial polypeptides and sequences having 70% or 90% identity to
XX the nucleotide and protein sequences. The nucleotides and polypeptides
XX are useful as diagnostic markers or therapeutic target for the diseases
XX or morbid states. They are also useful for treating osteoporosis,
XX neurological diseases, Alzheimer's diseases, Parkinson's diseases,
XX dementia and various cancers. This sequence corresponds to a nucleotide
XX sequence of the invention.
XX
XX Sequence 3660 BP; 887 A; 1049 C; 837 G; 887 T; 0 U; 0 Other;
XX
XX Query Match 42.6%; Score 212.6; DB 12; Length 3660;
XX Best Local Similarity 78.6%; Pred. No. 5.5e-56;
XX Matches 254; Conservative 0; Mismatches 69; Indels 0; Gaps 0;
XX
XX 37 TGGGACCTGGGATGGAGTTAGCGCTCAGTTAATAGTAATCTCAGCTGGGTGGCGTGGC 96
XX 1862 TGTGGCTTGGCTGGCATCAGACAGGCTTTAAGAGGAGATGTCTCAGCTGGGTGGCGTGGC 1803
XX
XX 97 TCATCTCTGTATTTCCAGCACTTTGGGAGACCGAGTTGGTGGATCACTTGAGAGCAGGA 156
XX 1802 TCACACCTGTAATTTAGCACTTTGGGAGGCGGAGTGGTGGATCCCTTGAGGTGAGGA 1743
XX
XX 157 GTTTCAGACACGCTGGGCAACATGGCAAAACACTATCTCTAATAAATAAATAAATAAATA 216
XX 1742 GTTCGAGACACGCTGGGCAACATGGCAAAACCCCTCTCTATGAAAAATAGAAAAATA 1683
XX
XX 217 GCCAGCTGTGGTGGCACTTTGCTATAGTCCAGCTACACAGGAGCTGGGGCAGAGGAAT 276
XX 1682 GCCAGGCGTGGTGGCAGGCGCTGTAAATTCAGCTACTTCAGGAGCTGAGGCACAAGAT 1623
XX
XX 277 CACTTGAACCTGGGAGGTGGAGGTTGCGATGAGCCAAAGATTGCACCTGCACTTCAGGCC 336

Db 1622 CACTTGAACCCGGAGGACAGGTTGCATATGAGCCAGACTGCGCACTGCACTCCAGCC 1563
QY 337 TCGAAAAAAGGTTAATTATTA 359
||| | | | | | | | | |
Db 1562 TGGGCGACAGAGTGAGTCTCTTA 1540

RESULT 13
ID AAK90931
XX AAK90931 standard; DNA; 10252 BP.
AC AAK90931;
XX
DT 05-NOV-2001 (first entry)
XX
DE Human digestive system antigen genomic sequence SEQ ID NO: 4507.
XX
KW Human; digestive system antigen; gene therapy; cancer; appendicitis;
KW ulcerative colitis; infection; Hirschsprung's disease; chronic colitis;
KW digestive system disorder; Meckel's diverticulum; ds.
XX
OS Homo sapiens.
XX
XX WO200155314-A2.
XX
PD 02-AUG-2001.
XX
PF 17-JAN-2001; 2001WO-US001324.
XX
XX 31-JAN-2000; 2000US-0179065P.
PR 04-FEB-2000; 2000US-0180628P.
PR 24-FEB-2000; 2000US-0184664P.
PR 02-MAR-2000; 2000US-0186350P.
PR 16-MAR-2000; 2000US-0198747P.
PR 17-MAR-2000; 2000US-0190076P.
PR 18-APR-2000; 2000US-0198123P.
PR 19-MAY-2000; 2000US-0205515P.
PR 07-JUN-2000; 2000US-0209467P.
PR 28-JUN-2000; 2000US-0214886P.
PR 30-JUN-2000; 2000US-0215135P.
PR 07-JUL-2000; 2000US-0216647P.
PR 07-JUL-2000; 2000US-0216880P.
PR 11-JUL-2000; 2000US-0217487P.
PR 11-JUL-2000; 2000US-0217496P.
PR 14-JUL-2000; 2000US-0218290P.
PR 26-JUL-2000; 2000US-0220963P.
PR 26-JUL-2000; 2000US-0220964P.
PR 14-AUG-2000; 2000US-0224518P.
PR 14-AUG-2000; 2000US-0224519P.
PR 14-AUG-2000; 2000US-0225213P.
PR 14-AUG-2000; 2000US-0225214P.
PR 14-AUG-2000; 2000US-0225266P.
PR 14-AUG-2000; 2000US-0225267P.
PR 14-AUG-2000; 2000US-0225268P.
PR 14-AUG-2000; 2000US-0225270P.
PR 14-AUG-2000; 2000US-0225447P.
PR 14-AUG-2000; 2000US-0225757P.
PR 14-AUG-2000; 2000US-0225758P.
PR 14-AUG-2000; 2000US-0225759P.
PR 18-AUG-2000; 2000US-0226279P.
PR 22-AUG-2000; 2000US-0226681P.
PR 22-AUG-2000; 2000US-0226686P.
PR 22-AUG-2000; 2000US-0227182P.
PR 23-AUG-2000; 2000US-0227009P.
PR 30-AUG-2000; 2000US-0228924P.
PR 01-SEP-2000; 2000US-0229287P.
PR 01-SEP-2000; 2000US-0229343P.
PR 01-SEP-2000; 2000US-0229344P.
PR 01-SEP-2000; 2000US-0229345P.
PR 05-SEP-2000; 2000US-0229509P.
PR 05-SEP-2000; 2000US-0229513P.
PR 06-SEP-2000; 2000US-0230437P.
PR 06-SEP-2000; 2000US-0230438P.
PR 08-SEP-2000; 2000US-0231242P.
PR 08-SEP-2000; 2000US-0231243P.
PR 08-SEP-2000; 2000US-0231244P.
PR 08-SEP-2000; 2000US-0231413P.
PR 08-SEP-2000; 2000US-0231414P.
PR 08-SEP-2000; 2000US-0232080P.
PR 08-SEP-2000; 2000US-0232081P.
PR 12-SEP-2000; 2000US-0231968P.
PR 14-SEP-2000; 2000US-0232397P.
PR 14-SEP-2000; 2000US-0232398P.
PR 14-SEP-2000; 2000US-0232399P.
PR 14-SEP-2000; 2000US-0232400P.
PR 14-SEP-2000; 2000US-0232401P.
PR 14-SEP-2000; 2000US-0233063P.
PR 14-SEP-2000; 2000US-0233064P.
PR 14-SEP-2000; 2000US-0233065P.
PR 21-SEP-2000; 2000US-0234223P.
PR 21-SEP-2000; 2000US-0234274P.
PR 25-SEP-2000; 2000US-0234997P.
PR 25-SEP-2000; 2000US-0234998P.
PR 26-SEP-2000; 2000US-0235484P.
PR 27-SEP-2000; 2000US-0235834P.
PR 27-SEP-2000; 2000US-0235836P.
PR 29-SEP-2000; 2000US-0236327P.
PR 29-SEP-2000; 2000US-0236367P.
PR 29-SEP-2000; 2000US-0236368P.
PR 29-SEP-2000; 2000US-0236369P.
PR 29-SEP-2000; 2000US-0236370P.
PR 02-OCT-2000; 2000US-0236802P.
PR 02-OCT-2000; 2000US-0237037P.
PR 02-OCT-2000; 2000US-0237038P.
PR 02-OCT-2000; 2000US-0237039P.
PR 02-OCT-2000; 2000US-0237040P.
PR 13-OCT-2000; 2000US-0239355P.
PR 13-OCT-2000; 2000US-0239937P.
PR 20-OCT-2000; 2000US-0240960P.
PR 20-OCT-2000; 2000US-0241221P.
PR 20-OCT-2000; 2000US-0241785P.
PR 20-OCT-2000; 2000US-0241786P.
PR 20-OCT-2000; 2000US-0241787P.
PR 20-OCT-2000; 2000US-0241808P.
PR 20-OCT-2000; 2000US-0241809P.
PR 20-OCT-2000; 2000US-0241826P.
PR 01-NOV-2000; 2000US-024617P.
PR 08-NOV-2000; 2000US-0246474P.
PR 08-NOV-2000; 2000US-0246475P.
PR 08-NOV-2000; 2000US-0246476P.
PR 08-NOV-2000; 2000US-0246477P.
PR 08-NOV-2000; 2000US-0246478P.
PR 08-NOV-2000; 2000US-0246523P.
PR 08-NOV-2000; 2000US-0246528P.
PR 08-NOV-2000; 2000US-0246529P.
PR 08-NOV-2000; 2000US-0246524P.
PR 08-NOV-2000; 2000US-0246525P.
PR 08-NOV-2000; 2000US-0246526P.
PR 08-NOV-2000; 2000US-0246527P.
PR 08-NOV-2000; 2000US-0246528P.
PR 08-NOV-2000; 2000US-0246532P.
PR 08-NOV-2000; 2000US-0246609P.
PR 08-NOV-2000; 2000US-0246610P.
PR 08-NOV-2000; 2000US-0246611P.
PR 08-NOV-2000; 2000US-0246613P.
PR 17-NOV-2000; 2000US-0249207P.
PR 17-NOV-2000; 2000US-0249208P.
PR 17-NOV-2000; 2000US-0249209P.
PR 17-NOV-2000; 2000US-0249210P.
PR 17-NOV-2000; 2000US-0249211P.
PR 17-NOV-2000; 2000US-0249212P.
PR 17-NOV-2000; 2000US-0249213P.
PR 17-NOV-2000; 2000US-0249214P.
PR 17-NOV-2000; 2000US-0249215P.
PR 17-NOV-2000; 2000US-0249216P.
PR 17-NOV-2000; 2000US-0249217P.
PR 17-NOV-2000; 2000US-0249218P.
PR 17-NOV-2000; 2000US-0249244P.

PR	17-NOV-2000;	2000US-0249245P.	
PR	17-NOV-2000;	2000US-0249264P.	
PR	17-NOV-2000;	2000US-0249265P.	
PR	17-NOV-2000;	2000US-0249297P.	
PR	17-NOV-2000;	2000US-0249299P.	
PR	17-NOV-2000;	2000US-0249300P.	
PR	01-DEC-2000;	2000US-0250160P.	
PR	01-DEC-2000;	2000US-0250391P.	
PR	05-DEC-2000;	2000US-0251030P.	
PR	05-DEC-2000;	2000US-0251988P.	
PR	05-DEC-2000;	2000US-0256719P.	
PR	06-DEC-2000;	2000US-0251479P.	
PR	08-DEC-2000;	2000US-0251856P.	
PR	08-DEC-2000;	2000US-0251868P.	
PR	08-DEC-2000;	2000US-0251869P.	
PR	08-DEC-2000;	2000US-0251989P.	
PR	08-DEC-2000;	2000US-0251990P.	
PR	11-DEC-2000;	2000US-0254097P.	
PR	05-JAN-2001;	2001US-0259678P.	
XX	(HUMA-)	HUMAN GENOME SCI INC.	
XX	Rosen CA, Barash SC, Ruben SM;		
XX	WPI; 2001-502630/55.		
XX	Polynucleotides encoding digestive system antigens, useful for		
PT	diagnosing, treating, preventing and/or prognosing disorders of the		
PT	digestive system, particularly cancer and cancer metastases.		
XX	Disclosure; SEQ ID NO 4507; 986pp; English.		
CC	The present invention provides the protein and coding sequences of a		
CC	number of human digestive system antigens. These can be used in the		
CC	diagnosis, treatment and prevention of digestive system disorders,		
CC	including cancer, Meckel's diverticulum, bacterial or parasitic		
CC	infections, appendicitis, Hirschsprung's disease, chronic colitis or		
CC	ulcerative colitis. The present sequence is a genomic DNA fragment		
CC	encoding a digestive system antigen of the invention		
XX	Sequence 10252 BP; 2689 A; 2608 C; 2537 G; 2418 T; 0 U; 0 Other;		
SQ	Query Match 42.4%; Score 211.8; DB 4; Length 10252;		
	Best Local Similarity 80.1%; Pred. No. 1.5e-55;		
	Matches 249; Conservative 0; Mismatches 62; Indels 0; Gaps 0;		
QY	64 GTTAATAGTAACCTATTAGCCAGGTGCGGTGGCTCATGTCCTATTCCCGACCTTTGGG 123		
Db	8374 GTTAATAAAGAGGTGGGCGCAGGACGTGGCTCATGCTGTAAATCCCGACCTTTGGG 8433		
QY	124 AGACCCAGTGGGTGGATCACTTGAGCAGGAGTTTGAGACCAGCTGGCCCAACATGGC 183		
Db	8434 AGTCTGAGGCGGTGGATCACTGAGTTCGAGAGTTTCGAGACCGCTGGCCCAACATGAC 8493		
QY	184 AAAACACTATCTCTAATAAAAATACAAAATTAAGCCAGGTGGTGGCCTTGCCCTATAG 243		
Db	8494 AAAACCCCATCTCTACTAAAAATACAAAATTAAGCTGGGTGGTGGCAGTTTCCCTATAA 8553		
QY	244 TCCAGCTACAGAGGCGTGGGCGAGAGAAATCACTTGAACCTGGAGGTGGAGGTTCG 303		
Db	8554 TCTCAGCTACTCAGGAGGCTGAGGCGAGGAAATCGCTTGAACCCAGGAGCGCGAGGTTCG 8613		
QY	304 AGTGAGCCAGATTGGACCTCACTCCAGCTCGGAAAAAAGGGTAAATTAATACTTT 363		
Db	8614 AGTGAGCCAGATTGGCCCACTGCATCCAGCTGGCAACAGAGCAAGATTCCTCTCA 8673		
QY	364 ACTTGAACCA 374		
Db	8674 AAACAAACAA 8684		
RESULT 14			
AAS31966			

ID	AAS31966 standard; DNA; 10252 BP.
XX	AAS31966;
AC	04-DEC-2001 (first entry)
XX	Human liver associated genomic DNA #140.
DT	Liver associated protein; human; mouse; rabbit; goat; horse; cat; dog;
XX	chicken; sheep; immunosuppressive; antiarthritic; vasotropic;
XX	antirheumatic; antiproliferative; cytostatic; cardiant; neuroprotective;
XX	cerebroprotective; nootropic; antibacterial; virucide; fungicide; cancer;
XX	ophthalmological; vulnery; gene therapy; autoimmune disease; neoplasia;
XX	hyperproliferative disorder; breast; liver; cardiovascular disorder; ds;
XX	cerebrovascular disorder; nervous system disorder; bacterial infection;
XX	fungal infection; viral infection; ocular disorder; endocrine disorder;
XX	gastrointestinal disorder; renal disorder; respiratory disorder;
XX	wound healing; skin aging; organ transplantation; tissue regeneration;
XX	anti-infertility.
OS	Homo sapiens.
XX	WO20015355-A1.
PN	02-AUG-2001.
XX	17-JAN-2001; 2001WO-US001351.
XX	31-JAN-2000; 2000US-0179065P.
XX	04-FEB-2000; 2000US-0180628P.
PR	24-FEB-2000; 2000US-0184664P.
PR	02-MAR-2000; 2000US-0186350P.
PR	16-MAR-2000; 2000US-0189874P.
PR	17-MAR-2000; 2000US-0190076P.
PR	18-APR-2000; 2000US-0198123P.
PR	19-MAY-2000; 2000US-0205515P.
PR	07-JUN-2000; 2000US-0209467P.
PR	28-JUN-2000; 2000US-0214886P.
PR	30-JUN-2000; 2000US-0215135P.
PR	07-JUL-2000; 2000US-0216647P.
PR	07-JUL-2000; 2000US-0216880P.
PR	11-JUL-2000; 2000US-0217487P.
PR	11-JUL-2000; 2000US-0217496P.
PR	14-JUL-2000; 2000US-0218290P.
PR	26-JUL-2000; 2000US-0220963P.
PR	26-JUL-2000; 2000US-0220964P.
PR	14-AUG-2000; 2000US-0224518P.
PR	14-AUG-2000; 2000US-0224519P.
PR	14-AUG-2000; 2000US-0225213P.
PR	14-AUG-2000; 2000US-0225214P.
PR	14-AUG-2000; 2000US-0225266P.
PR	14-AUG-2000; 2000US-0225267P.
PR	14-AUG-2000; 2000US-0225268P.
PR	14-AUG-2000; 2000US-0225270P.
PR	14-AUG-2000; 2000US-0225447P.
PR	14-AUG-2000; 2000US-0225757P.
PR	14-AUG-2000; 2000US-0225758P.
PR	14-AUG-2000; 2000US-0225759P.
PR	18-AUG-2000; 2000US-0226279P.
PR	22-AUG-2000; 2000US-0226681P.
PR	22-AUG-2000; 2000US-0226686P.
PR	22-AUG-2000; 2000US-0227182P.
PR	23-AUG-2000; 2000US-0227009P.
PR	30-AUG-2000; 2000US-0228924P.
PR	01-SEP-2000; 2000US-0229287P.
PR	01-SEP-2000; 2000US-0229343P.
PR	01-SEP-2000; 2000US-0229344P.
PR	01-SEP-2000; 2000US-0229345P.
PR	05-SEP-2000; 2000US-0229509P.
PR	05-SEP-2000; 2000US-0229513P.
PR	06-SEP-2000; 2000US-0230437P.
PR	06-SEP-2000; 2000US-0230438P.
PR	08-SEP-2000; 2000US-0231242P.


```

QY 184 AAAAAGCTATCTCTAATAAAAAATACAAAAATTAGCCAGGTGGTGGCACTTGCCCTATAG 243
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
8494 AAAACCCCATCTCTACTATAAAATACAAAAATTAGCTGGGTGGTGGCACTTGCCCTATAA 8553
QY 244 TCCCACTACACAGGAGGCTGGGGCAGAGAAATCACTTGAACTGGGAGGTGGAGTTGC 303
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
8554 TCTCAGCTACTCAGGAGGCTGAGGCAGGAGAAATCGCTTGAACCCAGGAGGCCGAGTTGC 8613
QY 304 AGTGAGCCCAAGATTGCAACCACTGCCTCCAGCTGGAAAAAAGGGTAATTAATTAACTTT 363
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
8614 AGTGAGCCCAAGATTGCGCCACTGCCTCCAGCTGGGCAACAGAGCAAGATTCCCATCTCA 8673
QY 364 ACTTGCAACCA 374
Db |||||
8674 AAACAAACAA 8684

```

Search completed: February 9, 2006, 13:50:51
Job time : 346.271 secs

THIS PAGE BLANK (USPTO)

GenCore version 5.1.7
Copyright (c) 1993 - 2006 Bioacceleration Ltd.

OM nucleic - nucleic search, using sw model

Run on: February 9, 2006, 13:34:47 ; Search time 2358.93 Seconds
(without alignments)
9917.025 Million cell updates/sec

Title: US-10-607-806-1_COPY_1_500

Perfect score: 499.6

Sequence: 1 Gacctactctgacctttgtg.....accacataagaggactga 500

Scoring table: IDENTITY_NUC

Gapop 10_0 , Gapext 1.0

Searched: 41078325 seqs, 23393541228 residues

Total number of hits satisfying chosen parameters: 82156650

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : EST:*

- 1: gb_est1:*
- 2: gb_est2:*
- 3: gb_est3:*
- 4: gb_hic:*
- 5: gb_est4:*
- 6: gb_est5:*
- 7: gb_est6:*
- 8: gb_est7:*
- 9: gb_gss1:*
- 10: gb_gss2:*
- 11: gb_gss3:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
C 1	212.4	42.5	917	5	BUI53359
C 2	210.8	42.2	687	11	CR958221
C 3	210.4	42.1	414	1	AW511559
4	209.8	42.0	719	9	AQ423663
5	209.6	42.0	683	10	AG119835
C 6	209.4	41.9	649	9	B71494
C 7	209.2	41.9	599	9	AQ586662
C 8	208.8	41.8	799	3	BM679026
9	208.4	41.7	455	9	AQ037381
10	206.8	41.4	374	8	F35684
C 11	206.6	41.4	573	5	BX470494
C 12	206.6	41.4	600	3	BM790237
C 13	206.6	41.4	657	3	BM788692
14	206.6	41.4	795	11	CR960649
C 15	206.4	41.3	514	2	BE169870
C 16	206.4	41.3	891	6	CD300637
17	206.2	41.3	666	9	AQ372692
18	206	41.2	769	5	BH854207
C 19	206	41.2	849	8	DN602222
C 20	205.6	41.2	416	1	A1446478
C 21	205.6	41.2	478	9	B67141
C 22	205.4	41.1	304	1	A1915081

C 23	205.4	41.1	570	5	BX491191	BX491191 DKFZp686B
C 24	205.4	41.1	584	6	CB128054	CB128054 K-EST0177
C 25	205.4	41.1	697	9	AQ054805	AQ054805 CIT-HSP-2
26	205.2	41.1	420	2	BE062545	BE062545 QV4-BT025
C 27	205.2	41.1	440	9	AQ112000	AQ112000 CIT-HSP-2
C 28	205.2	41.1	536	9	AQ480181	AQ480181 RPT-11-2
29	205.2	41.1	603	3	BP872932	BP872932 BP872932
C 30	205	41.0	437	1	AA644090	AA644090 ab62b04.8
C 31	205	41.0	550	9	AQ377547	AQ377547 RPT11-14
C 32	205	41.0	637	1	A1114494	A1114494 HA1076 Hu
C 33	205	41.0	4903	4	CR857861	CR857861 Pongc pyg
34	204.8	41.0	431	9	AQ268195	AQ268195 RPT11-73
C 35	204.8	41.0	546	9	AQ548438	AQ548438 RPT-11-4
C 36	204.8	41.0	755	5	BX644757	BX644757 DKFZp781B
C 37	204.8	41.0	5938	4	HSM807608	BX647463 Homo sapi
C 38	204.6	41.0	371	5	BUI752091	BUI752091 UI-1-BB0-
C 39	204.6	41.0	386	8	TS4464	TS4464 yb06b05.s1
C 40	204.6	41.0	563	5	BUI072494	BUI072494 im47h01.x
C 41	204.6	41.0	833	9	BZ599987	BZ599987 WHACUC62TR
42	204.6	41.0	910	2	BF337320	BF337320 602034695
43	204.4	40.9	681	2	BG618589	BG618589 602646014
C 44	204.2	40.9	467	2	BF940118	BF940118 nac58c10.
C 45	204.2	40.9	560	1	AL589348	AL589348 DKFZp451K

ALIGNMENTS

RESULT 1
BUI53359/c
LOCUS
DEFINITION AGENCOURT_7934963 NIH_MGC_68 Homo sapiens cDNA clone IMAGE:6009715
5' mRNA sequence.
ACCESSION BUI53359
VERSION BUI53359.1 GI:22666891
KEYWORDS EST.
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini;
Hominidae; Homo.
REFERENCE 1 (bases 1 to 917)
AUTHORS NIH-MGC <http://mgc.nci.nih.gov/>.
TITLE National Institutes of Health, Mammalian Gene Collection (MGC)
JOURNAL Unpublished (1999)
COMMENT Contact: Robert Strausberg, Ph.D.
Email: cgabbs-re@mail.nih.gov
Tissue Procurement: DCTD/DTF/Gazdar
CDNA Library Preparation: Life Technologies, Inc.
CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
DNA Sequencing by: Agencourt Bioscience Corporation
Clone distribution: MGC clone distribution information can be found through the I.M.A.G.E. Consortium/LLNL at:
<http://image.llnl.gov>
Plate: LLAM13196 row: j column: 20
High quality sequence stop: 529.
FEATURES
Location/Qualifiers
1..917
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/clone="IMAGE:6009715"
/tissue_type="large cell carcinoma"
/lab_host="DH10B (phage-resistant)"
/clone_lib="NIH_MGC_68"
/note="Organ: lung; Vector: pCMV-SPORT6; Site 1: NotI; Site 2: SalI; Cloned unidirectionally. Primer: Oligo dT. Average insert size 1.8 kb. Library constructed by Life Technologies."
ORIGIN
Query Match 42.5%; Score 212.4; DB 5; Length 917;
Best Local Similarity 80.3%; Pred. No. 3e-26;

Matches 249; Conservative 0; Mismatches 61; Indels 0; Gaps 0;

QY 67 AATAGTAATCATTTAGCCAGGTGGGTGGCTCATGTCTGTATATCCAGCACTTTGGGAGA 126
 |||||
 Db 575 AAAAAAAAAAAGTGGCGGTGGCTCATGTCTGTATATCCAGCACTTTGGGAGG 516
 |||||

QY 127 CCGAGTTGGGTGGATCATTTGAGACGAGAGTTTGAGACCGAGCTGGCCACATGGCAAA 186
 |||||
 Db 515 CCGAGTTGGGAGGATCATTTAAGCGCGAGGTTTGAGACCGAGCTGGCCAAAATGGCGAA 456
 |||||

QY 187 ACATATCTCTAATAAAAAATACAAAAATTTAGCCAGGTGGTGGGACATTTGCTATAGTCC 246
 |||||
 Db 455 ACCCATCTCTACTAAAAATCAAAAAATTTAGCCGGGCGTGGTGGCAGCGGCTGTATCC 396
 |||||

QY 247 CAGCTTACACAGGAGGCTGGGCGAGAGAATCATTTGAACTCTGGAGGTGGAGTTGCAGT 306
 |||||
 Db 395 CAGCTACTCAGGAGGCTGAGGCGAGAGAATCGCTGAAACCCAGGAGGTGGAGTTGCAGT 336
 |||||

QY 307 GAGCAAGATTGCACACTGCACTCAGCTGGAAGAAAAAGGTAATTAATACTTTACT 366
 |||||
 Db 335 GAGCAAGATTGCACACTGCACTCAGCTGGAAGAAAAAGGTAATTAATACTTTACT 276
 |||||

QY 367 TGCACACATA 376
 |||||

Db 275 ATAAAAATA 266
 |||||

RESULT 2

CR958221

LOCUS CR958221.1 687 bp DNA linear GSS 07-JUN-2005

DEFINITION Homo sapiens PAC end sequence of RZPDB737G0112D from genomic library (orig. Pieter J. de Jong library RPCI-1), genomic survey sequence.

ACCESSION CR958221

VERSION CR958221.1 GI:669555562

KEYWORDS GSS.

SOURCE Homo sapiens (human)

ORGANISM Homo sapiens

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini; Hominidae; Homo.

REFERENCE 1 (bases 1 to 687)

AUTHORS Schwarz, F., Neubert, P., Schneider, D., Peters, M. and Korn, B.

TITLE Direct Submission

JOURNAL Submitted (07-JUN-2005) RZPD Deutsches Ressourcenzentrum fuer Genomforschung GmbH, Im Neuenheimer Feld 515, D-69120 Heidelberg, Germany

COMMENT RZPD; RZPDB737G0112D; RPI-310P17; derived from Pieter J. de Jong library RPCI-1; http://www.rzpd.de/cgi-bin/products/cl.cgi?CloneID=RZPDB737G0112D RZPDLIB; (Human Genomic Set - RZPD 1.0) RZPD LIB No. 737 http://www.rzpd.de/cgi-bin/products/set.cgi?libNo=737 http://www.rzpd.de/products/genomicset/

Contact: Inge Arlt

RZPD Deutsches Ressourcenzentrum fuer Genomforschung GmbH, Heubnerweg 6, D-14059 Berlin, Germany

Tel: +49 30 32639 100

Fax: +49 30 32639 111

www.rzpd.de

This clone is available from RZPD; Contact RZPD (customer.service@rzpd.de) for further information. Clone distribution: http://www.rzpd.de/products/genomicset/

Seq-primer: T7 (TA-TAC-GAC-TCA-CTA-TAG-GG)

Class: BAC ends.

FEATURES

source Location/Qualifiers

1. .687

/organism="Homo sapiens"

/mol_type="genomic DNA"

/db_xref="taxon:9606"

/clone="RZPDB737G0112D"

/sex="Male"

/clone_lib="RPCI-1"

/note="Vector: pCYPAC2; RPCI-1 Human Male PAC Library"

ORIGIN

Query Match 42.2%; Score 210.8; DB 11; Length 687;
 Best Local Similarity 83.6%; Pred. No. 6e-26; Indels 0; Gaps 0;
 Matches 239; Conservative 0; Mismatches 47;

QY 67 AATAGTAATCATTTAGCCAGGTGGGTGGCTCATGTCTGTATATCCAGCACTTTGGGAGA 126
 |||||
 Db 108 AAGGGACAATTCAGCCAGCGCGAGTGGCTCAGCGCTGTAAATCCAGCACTTTGGGAGG 167
 |||||

QY 127 CCGAGTTGGGTGGATCATTTGAGACGAGAGTTTGAGACCGAGCTGGCCACATGGCAAA 186
 |||||
 Db 168 CCGAGTTGGGTGGATCATTTGAGCTCAGGAGTTTGAGACCGAGCTGGCCAACTGGTCAA 227
 |||||

QY 187 ACATATCTCTAATAAAAAATACAAAAATTTAGCCAGGTGGTGGGACATTTGCTATAGTCC 246
 |||||
 Db 228 ACCCATCTCTACTAAAAATACAAAAATTTAGTGGGTGGTGGCAGGTGGCTATATCC 287
 |||||

QY 247 CAGCTTACACAGGAGGCTGGGCGAGAGAATCATTTGAACTCTGGAGGTGGAGTTGCAGT 306
 |||||

Db 288 CAGCTACTTGGAGGCTGAGGCGAGAGAATCATTTGAACTCTGGAGGTGGAGTTTCAGT 347
 |||||

QY 307 GAGCAAGATTGCACACTGCACTCAGCTGGAAGAAAAAGGTTAA 352
 |||||

Db 348 GAGCGAGACTGTGCATTTGCATTTCCGGCCTGGGCAATAAGAGCAA 393
 |||||

RESULT 3

AW511559/c

LOCUS AW511559.1 414 bp mRNA linear EST 03-MAR-2000

DEFINITION xu58b07.x1 NCI CGAP Utl1 Homo sapiens cDNA clone IMAGE:2805877 3' similar to contains 'Alu repetitive element', mRNA sequence.

ACCESSION AW511559

VERSION AW511559.1 GI:7149561

KEYWORDS EST.

SOURCE Homo sapiens (human)

ORGANISM Homo sapiens

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini; Hominidae; Homo.

REFERENCE 1 (bases 1 to 414)

AUTHORS NCI-CGAP

TITLE NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.

JOURNAL National Cancer Institute, Cancer Genome Anatomy Project (CGAP), Tumor Gene Index

COMMENT Unpublished (1997)

Contact: Robert Strausberg, Ph.D.
 Email: cgapbs-remail.nih.gov

Tissue Procurement: Christopher Moskaluk, M.D., Ph.D., Michael R. Emmert-Buck, M.D., Ph.D.

cDNA Library Preparation: Life Technologies, Inc.

cDNA Library Arrayed by: Greg Lennon, Ph.D.

DNA Sequencing by: Washington University Genome Sequencing Center

Clone distribution: NCI-CGAP clone distribution information can be found through the I.M.A.G.E. Consortium/LLNL at: image.llnl.gov/image/html/iresources.shtml

Seq primer: -40UP from Gibco

High quality sequence stop: 411.

Location/Qualifiers

1. .414

/organism="Homo sapiens"

/mol_type="mRNA"

/db_xref="taxon:9606"

/clone="IMAGE:2805877"

/tissue_type="well-differentiated endometrial adenocarcinoma, 7 pooled tumors"

/lab_host="DH10B"

/clone_lib="NCI CGAP Utl1"

/note="Organ: uterus; Vector: pCMV-SPORT6; Site 1: SalI; Site 2: NotI; Cloned unidirectionally. Primer: Oligo dt. Average insert size 1.75 kb. Life technologies catalog #: 11538-014"

ORIGIN

```
Query Match      42.1%; Score 210.4; DB 1; Length 414;
Best Local Similarity 78.1%; Pred. No. 8.2e-26;
Matches 253; Conservative 0; Mismatches 71; Indels 0; Gaps 0;

QY 31 AGCATATGGGACCTGGGATGAGTTAGCGCTCAGTTAATAGTAACTCATTAGCCAGGTC 90
DB 341 AAGATTGAGAAGTCGGAGGAAGTATACATTTTAAGATCAGTTTCTGGCCGGTGC 282
QY 91 GGTGGCTCATGCTCTGATTTCCAGCACTTTGGGAGACCGAGTTGGGTGGATCATTGAGA 150
DB 281 GGTGGCTCAGCGCTGTAATCCAGCACTTTGGGAGCGGAGGTAGGCTGATCATTGAGG 222
QY 151 GCAGAGTTTGAGACCGCTGGCCCAACATGGCAAAACATCTCTTAATAAATAACAA 210
DB 221 TCAGGAGTCCAGACCGCTGGCCCAACATAGTAAACCCCATCTCTACTAAATAACAA 162
QY 211 AAATTAGCAGGTGGTGGCACTTGCCTATAGTCCAGCTACACAGGAGCTGGGCGAG 270
DB 161 AAATTAGCAGGATGGTGGTACATGCCGGTAGTCCCGAGTACTCAGGAGGCTGGGCGAG 102
QY 271 AAGATCACTTGAACCTGGGAGGTGGAGTTGCAGTGAGCCAAAGATTGCACCACTGCACT 330
DB 101 GAGATCGCTTGAACCCGGAGGCGCAGGTTGCAGTGAGCCAAAGATCGTGCATCTGCACT 42
QY 331 CCAGCTCGAAAAAAGGTAATT 354
DB 41 GCAGCTGGCAGCAAGTGAGACT 18

RESULT 4
LOCUS      AQ423663
DEFINITION CITBI-E1-2556C19.TF CITBI-E1 Homo sapiens genomic clone 2556C19,
ACCESSION  AQ423663
VERSION     AQ423663.1 GI:4481387
KEYWORDS    GSS.
SOURCE      Homo sapiens (human)
ORGANISM    Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini;
Hominidae; Homo.
REFERENCE   1 (bases 1 to 719)
AUTHORS    Zhao, S., Adams, M.D., Nierman, W., Malek, J., Shizuya, H., Simon, M. and
            Venter, J.C.
TITLE       Use of BAC End Sequences from CalTech Libraries for Sequence-Ready
JOURNAL     Map Building
COMMENT     Unpublished (1997)
            Other GSSs: CITBI-E1-2556C19.TR
            Contact: Shaying Zhao, William Nierman, Mark Adams
            Department of Eukaryotic Genomics
            The Institute for Genomic Research
            9712 Medical Center Dr., Rockville, MD 20850
            Tel: 301 838 0200
            Fax: 301 838 0208
            Email: hbe@tigr.org
            Clones are available from Research Genetics (info@resgen.com). BAC
            end search page:
            http://www.tigr.org/tdb/humgen/bac_end_search/bac_end_search.html.
            Seq primer: M13-21
            Class: BAC ends.

FEATURES             Location/Qualifiers
     source           1..719
                     /organism="Homo sapiens"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:9606"
                     /clone="2556C19"
                     /sex="male"
                     /cell_type="sperm"
                     /clone_lib="CITBI-E1"
                     /note="vector: pHELOBAC11; Site_1: EcoRI; Site_2: EcoRI;
                     CalTech Human BAC Library D"

ORIGIN
Query Match      42.0%; Score 209.6; DB 10; Length 683;

LOCUS      AG119835
DEFINITION Pan troglodytes DNA, clone: PTB-128G10.F, genomic survey sequence.
ACCESSION  AG119835
VERSION     AG119835.1 GI:16740354
KEYWORDS    GSS.
SOURCE      Pan troglodytes (chimpanzee)
ORGANISM    Pan troglodytes
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini;
Hominidae; Pan.
REFERENCE   1
AUTHORS     Fujiyama, A., Hattori, M., Toyoda, A., Taylor, T.D., Yada, T.,
            Totoki, Y., Watanabe, H. and Sakaki, Y.
TITLE       BAC end sequences of library PTB
JOURNAL     Unpublished
COMMENT     2 (bases 1 to 683)
            Fujiyama, A., Hattori, M., Toyoda, A., Taylor, T.D., Yada, T.,
            Totoki, Y., Watanabe, H. and Sakaki, Y.
            Submitted (02-AUG-2001) Asao Fujiyama, The Institute of Physical
            and Chemical Research (RIKEN), Genomic Sciences Center (GSC);
            1-7-22 Suehiro-chou, Tsurumi-ku, Yokohama, Kanagawa 230-0045, Japan
            (E-mail: chimpbes@gsc.riken.go.jp, URL: http://hgp.gsc.riken.go.jp/,
            Tel: 81-45-503-9111, Fax: 81-45-503-9170)
            Clones are derived from the chimpanzee BAC library PTB. This BAC end
            was generated during the R&D process and may have higher chance of
            clone tracking errors.
            PRIMERS
            Sequencing: -21M13
            LIBRARY
            Vector      : pKS145
            R.Site 1    : SacI
            R.Site 2    : SacI.
            Location/Qualifiers
     source           1..683
                     /organism="Pan troglodytes"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:9598"
                     /clone="PTB-128G10.F"
                     /sex="male"
                     /cell_type="lymphoblast"
                     /clone_lib="PTB Chimpanzee Male BAC Library"

ORIGIN
Query Match      42.0%; Score 209.6; DB 10; Length 683;
```



```

/note="Vector: pBACe3.6; Site 1: EcoRI; Site 2: EcoRI;
RPC111 Human Male BAC Library"

ORIGIN
Query Match      41.9%; Score 209.2; DB 9; Length 599;
Best Local Similarity 79.7%; Pred. No. 1.2e-25;
Matches 247; Conservative 0; Mismatches 63; Indels 0; Gaps 0;

QY 67 AATAGTAATCACTATTAGCCAGGTGCGGTGCTCATGTCTGTATTCACGACCTTTGGGAGA 126
Db 61 ATTAACATGTCACAGCGTGGTGCACTGCTCATGCTGTATTCACGACCTTTAGGAGG 120

QY 127 CCGAGTTGGTGGATCACTTGAGAGAGAGAGTTGAGACCACTGGCCCAACATGCAAA 186
Db 121 CCGAGGTGAGTGGATCACTTGAGGCGAGAGTTCAAGACCACTGGCCCAACATGTA 180

QY 187 ACACTATCTCTAATAAATACAAAATTTAGCCAGGTGTTGGGCACTTGCCTATAGTCC 246
Db 181 ACCCGTCTTTACTAAAATTTCAAAAATTTAGCCAGGTGTTGGGCACTTGCCTATAGTCC 240

QY 247 CAGCTACACAGGAGGTGGGGCAGAGAAATCACTTCAACCTGGGAGGTGGAGTTGCACT 306
Db 241 CAGCTATTCAGCAGGCTGAGGCGAGGAGATCCCTTAACCTGGGAGGTGGAGTTGCACT 300

QY 307 GAGCAAGATTTCACCACTGCACTCAGCTCGAAGAAAAGGTTAATTAATCACTTTACT 366
Db 301 GAGGCAAGATTTCACCACTGCACTCAGCTCGAAGAAAAGGTTAATTAATCACTTTACT 360

QY 367 TGCACCACTA 376
Db 361 CACACACAAA 370

RESULT 8
BM679026/c
LOCUS
DEFINITION
UI-E-E00-ahx-j-02-0-UI.s1 UI-E-E00 Homo sapiens cdna clone
UI-E-E00-ahx-j-02-0-UI 3', mRNA sequence.
ACCESSION
BM679026
VERSION
BM679026.1 GI:18988922
KEYWORDS
EST.
SOURCE
Homo sapiens (human)
ORGANISM
Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini;
Hominidae; Homo.
1 (bases 1 to 799)
Bonaldo,M.F., Lennon,G. and Soares,M.B.
Normalization and subtraction: two approaches to facilitate gene
discovery
Genome Res. 6 (9), 791-806 (1996)
8889548
Contact: Soares, MB
Coordinated Laboratory for Computational Genomics
University of Iowa
375 Newton Road, 4156 MBRRF, Iowa City, IA 52242, USA
Tel: 319 335 8250
Fax: 319 335 9565
Email: bento-soares@uiowa.edu
Tissue Procurement: Dr. Gregg Hageman
cdna Library preparation: Dr. M. Bento Soares, University of Iowa
cdna Library Arrayed by: Dr. M. Bento Soares, University of Iowa
DNA Sequencing by: Dr. M. Bento Soares, University of Iowa
Clone Distribution: Researchers may obtain clones from Research
Genetics (www.reagen.com).
The following repetitive elements were found in this cdna
Sequence: 20-305, >ALU (matched complement)
Seq primer: M13 Forward
POLYA=Yes.
Location/Qualifiers
1..799
/organism="Homo sapiens"
/mol_type="mRNA"

FEATURES
Source

```

```

/db_xref="taxon:9606"
/clone="UI-E-E00-ahx-j-02-0-UI"
/tissue_type="fetal eye"
/dev_stages="fetal"
/lab_host="DH10B (Life Technologies) (T1 phage resistant)"
/clone_lib="UI-E-E00"
/note="Organ: eye; Vector: pT7T3-Pac (Pharmacia) with a
modified polylinker; Site 1: EcoRI; Site 2: Not I;
UI-E-E00 is a cdna library containing the following
tissue(s): fetal eye. The library was constructed
according to Bonaldo, Lennon and Soares, Genome Research,
6:791-806, 1996. First strand cdna synthesis was primed
with an oligo-dT primer containing a Not I site. Double
stranded cdna was ligated to an EcoRI adaptor, digested
with Not I, and cloned directionally into pT7T3-Pac
vector. The oligonucleotide used to prime the synthesis of
first-strand cdna contains a library tag sequence that is
located between the Not I site and the (dT)18 tail. The
sequence tag for this library is CGCGTATACC. This library
was created for the program, Gene Discovery in the Visual
System, supported by National Eye Institute (NEI).
TAG_TISSUE=human fetal eye
TAG_LIB=UI-E-E00
TAG_SEQ=CGCGTATACC"

```

ORIGIN

```

Query Match      41.8%; Score 208.8; DB 3; Length 799;
Best Local Similarity 77.8%; Pred. No. 1.2e-25;
Matches 252; Conservative 0; Mismatches 72; Indels 0; Gaps 0;

QY 31 AGCATATGGGACCTGGGATGGATTAGCGCTCACTTAATAGTAACCTCACTTAGCCAGGTGC 90
Db 355 AACATTTGAGAAGTCGAGAGAAAGTATTACATTTTAAGATCACTTTCCTGCGCGGTGC 296

QY 91 GGTGGCTCATGTCGTATATCCAGCACTTTGGGAGACCGAGTTGGGTGGATCACTTGAGA 150
Db 295 GGTGGCTCACGCTCTAATCCAGCACTTTGGGAGCGCGAGTTAGGCTGATCACTTGAGG 236

QY 151 GCAGGAGTTTGAGACCACTGGCCCAACATAGTAAACCCCATCTCTACTAAAAATACAA 210
Db 235 TCAGGAGCTCAGACCACTGGCCCAACATAGTAAACCCCATCTCTACTAAAAATACAA 176

QY 211 AAATTAGCCAGGTGTGGTGCACTTGGCTATAGTCCAGCTTACACAGAGGCTGGGCGAG 270
Db 175 AAATTAGCCAGGATGTGTGTACATGCGGTAGTCCAGCTACTCAGAGGCTGGGCGAT 116

QY 271 AGAATCACTTGAACCTGGGAGGTGGAGTTGCAGTGCAGCAAGATTGCACCTGCACT 330
Db 115 GAGAATCGCTTGAACCGCGAGGCGCAGTTGCAGTGAGCCAGATCGTGGCACTGCACT 56

QY 331 CCAGCCTGGAAGAAAAGGTAATT 354
Db 55 GCAGCCTGGCGACAAAGTGAGACT 32

```

```

RESULT 9
AQ037381
LOCUS
DEFINITION
CIT-HSP-2335L20.TR CIT-HSP Homo sapiens genomic clone 2335L20,
genomic survey sequence.
ACCESSION
AQ037381
VERSION
AQ037381.1 GI:3303213
KEYWORDS
GSS.
SOURCE
Homo sapiens (human)
ORGANISM
Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini;
Hominidae; Homo.
1 (bases 1 to 455)
Adams,M.D., Rounsley,S.D., Zhao,S., Field,C.E., Bass,S., Linher,K.,
Golden,K., Berry,K., Granger,D., Suh,E., Wible,C., Shizuya,H.,
Simon,M. and Venter,J.C.
Use of a random BAC End Sequence Database for Sequence-Ready Map
TITLE

```

JOURNAL COMMENT
 Building (1998)
 Unpublished (1998)
 Other GSSs: CIT-HSP-2335L20.TF
 Contact: Mark Adams
 Department of Eukaryotic Genomics
 The Institute for Genomic Research
 9712 Medical Center Dr., Rockville, MD 20850, USA
 Tel: 301 838 0200
 Fax: 301 838 0208
 Email: mdadams@tigr.org
 Clones are available from Research Genetics (info@resgen.com). BAC
 end search page:
http://www.tigr.org/tldb/hungen/bac_end_search/bac_end_search.html.
 Seq primer: M13 Reverse
 Class: BAC ends.

FEATURES
 source

Location/Qualifiers
 1..455
 /organism="Homo sapiens"
 /mol_type="genomic DNA"
 /db_xref="taxon:9606"
 /clone="2335L20"
 /sex="Male"
 /cell_type="Sperm"
 /clone_lib="CIT-HSP"
 /note="Vector: pBelobAC11; Site_1: HindIII; Site_2:
 HindIII"

ORIGIN

Query Match 41.7%; Score 208.4; DB 9; Length 455;
 Best Local Similarity 83.7%; Pred. No. 1.7e-25;
 Matches 236; Conservative 0; Mismatches 46; Indels 0; Gaps 0;
 QY 68 ATAGTAACCTATTAGCCAGGTGGGTGCTCATGTCTGTTATCCAGCACTTTGGGAGAC 127
 Db 124 AGACTCACATAGCCAGGCGAGTGGCTCATGCTGTAATCCAGTACTTTGGGAGC 183
 QY 128 CGAGTTGGGTGATCACTTGAGAGCAGGAGTTTGAGACAGCCTGGCCCAACATGGCAAA 187
 Db 184 CGAGGCGGTGGATCACTTGAGTCAAGGAGTTCAAGACAGCCTGGCCCAACATGGGGAAG 243
 QY 188 CACTATCTCTATAAATAAATAAATAAATAGCCAGGTGGTGGCACTTGCCTATATGCTCC 247
 Db 244 CCCTGTCTCTACTATAAATAAATAAATAAATAGCCAGGCTGGTGGCGGCGCTGTAATCCC 303
 QY 248 AGCTACACAGGAGGCTGGGCGCAGAGAATCACTTTGAACCTGGAGGTGGAGGTTCAGTG 307
 Db 304 AGTACTCAGGAGGCTGAGCAGAGAGATCGCTTGAACCTGGAGGCGGAGGTTCAGTG 363
 QY 308 AGCCAAAGATGACCACTGCACTCCAGCTCGAAAAAAGGG 349
 Db 364 AGCCGAGATGTCCTACACTCCAGCTCGGCTGCTACAAGAG 405

RESULT 10
 F35684
 LOCUS HSPD32531 HM3 Homo sapiens cDNA clone sh5-000005-0/D10, mRNA
 DEFINITION sequence.
 F35684
 F35684.1 GI:4821310
 VERSION EST.
 KEYWORDS Homo sapiens (human)
 SOURCE Homo sapiens
 ORGANISM Eukaryota; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini;
 Homnidae; Homo.
 REFERENCE 1 (bases 1 to 374)
 AUTHORS Pandolfi, D., Topo, S., Trevisan, S., Scarso, S. and Valle, G.
 TITLE Identification of 4370 expressed sequence tags from a
 3'-end-specific cDNA library of human skeletal muscle by DNA
 sequencing and filter hybridization
 JOURNAL Genome Res. 6 (1), 35-42 (1996)

PUBMED COMMENT

8681137
 Contact: Valle G.
 CRIBI Biotechnology Centre
 University of Padua
 Via Trieste 75, 35121 Padua, Italy
 ABI Chromatograms and other information are available on WWW at
<http://grup.bio.unipd.it>.
 Location/Qualifiers
 1..374
 /organism="Homo sapiens"
 /mol_type="mRNA"
 /db_xref="taxon:9606"
 /clone="sh5-000005-0/D10"
 /sex="female"
 /tissue_type="pectoral muscle (after mastectomy)"
 /clone_lib="HM3"
 /note="Vector: pCDNAII (Invitrogen); Site 1: BstXI;
 Site 2: NotI; The library was constructed by G.
 Lanfranchi. This library is not subcloned nor normalized.
 The first strand cDNA was primed with a biotinylated
 oligo-dT-NotI primer
 (5'-biotin-AACCCGCTCGAGCGCGCTTTTCTTTT-3'). The
 ds cDNA was sonicated and size-selected in the range
 350-550 bp. The 3' specific fragments were selected by
 streptavidin coated magnetic beads, ligated to
 non-palindromic BstXI adapters, NotI digested and
 directionally cloned into BstXI-NotI cut pCDNAII vector."

FEATURES
 source

Location/Qualifiers
 1..374
 /organism="Homo sapiens"
 /mol_type="mRNA"
 /db_xref="taxon:9606"
 /clone="sh5-000005-0/D10"
 /sex="female"
 /tissue_type="pectoral muscle (after mastectomy)"
 /clone_lib="HM3"
 /note="Vector: pCDNAII (Invitrogen); Site 1: BstXI;
 Site 2: NotI; The library was constructed by G.
 Lanfranchi. This library is not subcloned nor normalized.
 The first strand cDNA was primed with a biotinylated
 oligo-dT-NotI primer
 (5'-biotin-AACCCGCTCGAGCGCGCTTTTCTTTT-3'). The
 ds cDNA was sonicated and size-selected in the range
 350-550 bp. The 3' specific fragments were selected by
 streptavidin coated magnetic beads, ligated to
 non-palindromic BstXI adapters, NotI digested and
 directionally cloned into BstXI-NotI cut pCDNAII vector."

ORIGIN

Query Match 41.4%; Score 206.8; DB 8; Length 374;
 Best Local Similarity 78.7%; Pred. No. 3.4e-25;
 Matches 247; Conservative 0; Mismatches 67; Indels 0; Gaps 0;
 QY 36 ATGGACCTGGATGGAGTTAGCGCTCAGTTAATAGTAACCTATAGCCAGGTGGGTGG 95
 Db 51 ATTAGACAGAGACTGAGCATCAATTTTAAAGTGAAGATGATGAGCCAGGTGGTGG 110
 QY 96 CTCATGCTGTATTCCTCCAGCACTTTGGGAGACCCAGATTGGGTGGATCACTTGGAGCAGG 155
 Db 111 CTCAGCTCTGTAATCCCGAGGTGGGAGGCTGAGGCGGTGGATCACTGAGGTGAGG 170
 QY 156 AGTTTGAGACAGCCTGGCCCAACATGGCAAAACACTATCTCTATAAATAAATAAATAAT 215
 Db 171 AGTTCAAGACAGCCTGGCCCAACATGGTGAACCCCATCTCTATAAATAAATAAATAAT 230
 QY 216 AGCCAGGTGGTGGCACTTGCCTATAGTCCAGTACACAGAGGCTGGGCGCAGAGAA 275
 Db 231 AGCCAGGCGGTGGTGGCAGTGTCTTGAATCCAGCTACTCTGGAGGCTGAGGCGAGAGAA 290
 QY 276 TCACCTTGAACCTGGGAGGTGGAGGTGGAGTGGAGCAAGATTGGACCACTGCACCTCCAGC 335
 Db 291 TTGCTTGAACCCAGGAGGAGGATTACAGTGAGCCGAGATGGCACCATTGCACCTAGC 350
 QY 336 CTGGAAAAAAGGG 349
 Db 351 CTGGGCACACAGAG 364

RESULT 11
 BX470494/c

LOCUS BX470494.1
 DEFINITION DKF2p686J21109 s1 686 (synonym: hlcc3) Homo sapiens cDNA clone
 sequence.
 BX470494
 BX470494.1 GI:31664821
 VERSION EST.
 KEYWORDS Homo sapiens (human)
 SOURCE Homo sapiens
 ORGANISM Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini;
 Homnidae; Homo.
 REFERENCE 1 (bases 1 to 573)
 AUTHORS Bahr, A., Lauber, J., Mewes, H.W., Weil, B., Amid, C., Osanger, A.,

Fobo, G., Han, M. and Wiemann, S.
EST (Bahr, A., Lauber, J., Mewes, H.W., Weil, B., et al.)
Unpublished (2003)
Contact: MIPS

Ingolstaedter Landstr. 1, D-85764 Neuherberg, Germany
This is the 3' sequence of the clone insert
Cloned from S. Wiemann, Molecular Genome Analysis, German Cancer
Research Center (DKFZ); Email s.wiemann@dkfz-heidelberg.de;
sequenced by Qiagen (Hilden/Germany) within the cDNA sequencing
consortium of the German Genome Project.

11 sequence also available.
This clone (DKFZp686J21109) is available at the RZPD in Berlin.
Please contact the RZPD: Ressourcenzentrum, Heubnerweg 6, 14059
Berlin-Charlottenburg, GERMANY; Email: clone@rzpd.de.

FEATURES

source
1..573
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/clone="DKFZp686J21109"
/dev_stage="adult"
/lab_host="DH10B"
/clone_lib="686 (synonym: hlcc3)"
/note="Vector: pTriplex2; Site_1: sf11A; Site_2: sf11B;
cDNA-collection"

ORIGIN

```

Query Match      41.4%; Score 206.6; DB 5; Length 573;
Best Local Similarity 84.1%; Pred. No. 3.2e-25;
Matches 233; Conservative 0; Mismatches 44; Indels 0; Gaps 0;

QY 76 TCATTAGCAGGTGGCGTCTGATTCATCCAGCAGCTTTGGAGACCGAGTTGG 135
   |||||
Db 303 TAATTGGCCAGCGGTGGCTCAGCCTGTAATCCAGCAGCTTTGGAGCCGAGGTG 244

QY 136 GTGATCATCTTGAGAGCAGGAGTTGAGACGAGCTGGCCAAACATGGCAAAACATATCT 195
   |||||
Db 243 ATGGATCACTTGAGACCGCGAGTTTGAGACCGAGCTGACCAACATGGCAAAACCCATCT 184

QY 196 CTAATAAAATACAAAATAGCCAGGTGGTGGCAGCTTGCCTATAGTCCAGCTACAC 255
   |||||
Db 183 CTACTAAAATACAAAATTTGGCGGCGATGGTGCCACATGCTGTAATCCAGCTACTC 124

QY 256 AGGAGGCTGGGCGAGAGAAATCACTTGAACCTGGGAGGTGGAGGTTCAGTGAGCCCAAGA 315
   |||||
Db 123 AGGAGGCTGAGCAGGAGAAATGCTGCTACTCTGGGAGGTGGAGGTTCGTGTGAGCCCAAGA 64

QY 316 TTGCACCACTGCATCCAGCTGGAAAGGTTAA 352
   |||||
Db 63 TCGCACCATTGCACTCTAGCTGGCGCAATACAGCA 27

```

RESULT 12
BM790237/c
LOCUS K-EST0070027 S22SNU16 Homo sapiens cDNA clone S22SNU16-4-C11 5',
DEFINITION mRNA sequence.

ACCESSION BM790237.1 GI:19138469

VERSION EST.

SOURCE Homo sapiens (human)

ORGANISM Homo sapiens
Eukaryota; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini;
Hominidae; Homo.

REFERENCE 1 (bases 1 to 600)
AUTHORS Kim, N.S., Hahn, Y., Oh, J.H., Lee, J.Y., Ahn, H.Y., Chu, M.Y., Kim, M.R.,
Oh, K.J., Cheong, J.E., Sohn, H.Y., Kim, J.M., Park, H.S., Kim, S. and
Kim, Y.S.

TITLE 21C Frontier Korean EST Project 2001
JOURNAL Unpublished (2002)
COMMENT Contact: Kim YS

Genome Research Center
Korea Research Institute of Bioscience & Biotechnology
52 Eoeun-dong Yuseong-gu, Daejeon 305-333, South Korea
Tel: +82-42-860-4470
Fax: +82-42-860-4409
Email: yongsung@mail.kribb.re.kr
Plate: 4 row: C column: 11
High quality sequence stop: 600.

FEATURES

source
1..600
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/clone="S22SNU16-4-C11"
/sex="F"
/tissue_type="Ascites"
/cell_type="lymphoblast-like"
/cell_line="SNU-16"
/lab_host="DH10B"
/clone_lib="S22SNU16"
/note="Organ: Stomach; Vector: pT7T3-Pac; Site_1: EcoRI;
Site_2: NotI; The S22SNU16 library was contributed by the
Soares laboratory and it was constructed as described by
Ronald, M.F., Lennon, G. and Soares, M.B. (1996), Genome
Research 6(9): 791-806. RNA was prepared from harvested
cells of SNU-16 culture. SNU-16 cell was obtained from
Korean Cell Line Bank (KCLB). SNU-16 was established from
ascitic fluids of Korean patients by Park J.G. et al.
(1990), Cancer Res 50: 2773-2780."

ORIGIN

```

Query Match      41.4%; Score 206.6; DB 3; Length 600;
Best Local Similarity 87.0%; Pred. No. 3.2e-25;
Matches 227; Conservative 0; Mismatches 34; Indels 0; Gaps 0;

QY 79 TTAGCCAGGTGGCGTCTGATTCATCCAGCAGCTTTGGAGACCGAGTTGGGTG 138
   |||||
Db 494 TCAGCCAGGCGAGTGGCTCAGCCTGTAATCCAGCAGCTTTGGAGCGGAGTGGGTG 435

QY 139 GATCAGCTGAGACGAGGTTTGAGACCGAGCTGGCCACATGGCAAAACATCTCTTA 198
   |||||
Db 434 GATCAGCTGAGGTCAGGAGTTTGAGACCGAGCTGGCCACATGGCAAAACATCTCTTA 375

QY 199 ATAAAATACAAAATTTAGCCAGGTGGTGGCAGCTTGCCTATAGTCCAGCTACACAGG 258
   |||||
Db 374 CTAATACTACAAAATTTAGCCAGGCTGGTGGCATGTGCTCTAACCCAGCTACTTGGG 315

QY 259 AGGCTGGGCGAGAGAAATCACTTGAACCTGGGAGGTGGAGGTTCGAGTGAGCCAAAGATTG 318
   |||||
Db 314 AGGCTGAGGCGAGAAATCACTTGAACCCAGGAGGTGGAGGTTCGAGTGAGCCAAAGATTG 255

QY 319 CACCACTGCCTCCAGCCTGG 339
   |||||
Db 254 TGCCACTGTACTCCAGCCTGG 234

```

RESULT 13

BM788692/c
LOCUS K-EST0067875 S10SNU1 Homo sapiens cDNA clone S10SNU1-20-H06 5',
DEFINITION mRNA sequence.

ACCESSION BM788692.1 GI:19136924

VERSION EST.

KEYWORDS Homo sapiens (human)

SOURCE Homo sapiens

ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini;
Hominidae; Homo.

REFERENCE 1 (bases 1 to 657)

AUTHORS Kim, N.S., Hahn, Y., Oh, J.H., Lee, J.Y., Ahn, H.Y., Chu, M.Y., Kim, M.R.,
Oh, K.J., Cheong, J.E., Sohn, H.Y., Kim, J.M., Park, H.S., Kim, S. and
Kim, Y.S.

<p>TITLE 21C Frontier Korean EST Project 2001 JOURNAL Unpublished (2002) COMMENT Contact: Kim YS Genome Research Center Korea Research Institute of Bioscience & Biotechnology 52 Boeun-dong Yuseong-gu, Daejeon 305-333, South Korea Tel: +82-42-860-4470 Fax: +82-42-860-4409 Email: yongsung@mail.kribb.re.kr Plate: 20 row: H column: 06 High quality sequence stop: 657.</p>		<p>KEYWORDS Homo sapiens (human) SOURCE Homo sapiens ORGANISM Homo sapiens Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini; Hominiidae; Homo. 1 (bases 1 to 795) Schwarz,F., Neubert,P., Schneider,P., Peters,M. and Korn,B. Direct Submission Submitted (05-JUN-2005) RZPD Deutsches Ressourcenzentrum fuer Genomforschung GmbH, Im Neuenheimer Feid 515, D-69120 Heidelberg, Germany RZPD; RZPDB737B062021D; RP11-56D10; derived from Pieter J. de Jong library RZPD-11; http://www.rzpd.de/cgi-bin/products/cl.cgi?Clonaid=RZPDB737B062021D RZPDB1B; (Human Genomic Set - RZPD 1.0) RZPD LIB No.737 http://www.rzpd.de/cgi-bin/products/set.cgi?libNo=737 http://www.rzpd.de/products/genomicset/ Contact: Inge Arian RZPD Deutsches Ressourcenzentrum fuer Genomforschung GmbH, Heubnerweg 6, D-14059 Berlin, Germany Tel: +49 30 32639 100 Fax: +49 30 32639 111 www.rzpd.de This clone is available from RZPD; Contact RZPD (customer.service@rzpd.de) for further information. Clone distribution: http://www.rzpd.de/products/genomicset/ Seq-primer: T7 (TAA-TAC-GAC-TCA-CTA-TAG-GG) Class: BAC ends.</p>		<p>REFERENCE 1 (bases 1 to 795) Schwarz,F., Neubert,P., Schneider,P., Peters,M. and Korn,B. Direct Submission Submitted (05-JUN-2005) RZPD Deutsches Ressourcenzentrum fuer Genomforschung GmbH, Im Neuenheimer Feid 515, D-69120 Heidelberg, Germany RZPD; RZPDB737B062021D; RP11-56D10; derived from Pieter J. de Jong library RZPD-11; http://www.rzpd.de/cgi-bin/products/cl.cgi?Clonaid=RZPDB737B062021D RZPDB1B; (Human Genomic Set - RZPD 1.0) RZPD LIB No.737 http://www.rzpd.de/cgi-bin/products/set.cgi?libNo=737 http://www.rzpd.de/products/genomicset/ Contact: Inge Arian RZPD Deutsches Ressourcenzentrum fuer Genomforschung GmbH, Heubnerweg 6, D-14059 Berlin, Germany Tel: +49 30 32639 100 Fax: +49 30 32639 111 www.rzpd.de This clone is available from RZPD; Contact RZPD (customer.service@rzpd.de) for further information. Clone distribution: http://www.rzpd.de/products/genomicset/ Seq-primer: T7 (TAA-TAC-GAC-TCA-CTA-TAG-GG) Class: BAC ends.</p>		<p>FEATURES Location/Qualifiers source 1..795 /organism="Homo sapiens" /mol_type="genomic DNA" /db_xref="taxon:9606" /clone="RZPDB737B062021D" /sex="Male" /clone_lib="RZPD-11" /note="Vector: pBACe3.6; RZPD-11 Human Male BAC Library"</p>		<p>ORIGIN Query Match 41.4%; Score 206.6; DB 11; Length 795; Best Local Similarity 82.8%; Pred. No. 2.9e-25; Matches 236; Conservative 0; Mismatches 49; Indels 0; Gaps 0; QY 68 ATAGTAATCTATAGCAGGTGGTGGCTCATGCTTATTTCCAGCACTTTGGGAGAC 127 Db 115 ATAAAAATTTAGGGCCAGCGCGTGGCTCATGCTATATATCCAGCACCTTTGGGAGC 174 QY 128 CGAGTTGGTGGATCAGTCTTGGAGAGGAGTTTGGAGACAGCTGGCCCAACATGGCAAAA 187 Db 175 CAAGCAGGTGGATCAGTCTTGGAGAGGAGTTTGGAGAGAGCTGGCCCAACATGGCAAAA 234 QY 188 CACTATCTCTAATAAAAAATACAAAAATTAGCAGGTGTGGTGGCACTTGCCTATAGTCCC 247 Db 235 CCCCATCTCTACCAAAAAATACAAAAATTAGCCAGGATGTGGTGGTGGCTTATAATCCC 294 QY 248 AGCTACACAGGAGCTGGGGCAGAGAAATCAGTCTGAACCTGGAGGTGGAGGTTCGAGTG 307 Db 295 AGCTACTCGGAGGCTGAACAGAGAGAAATCAGTCTGAACCCGGGAGGTGGAGGTTCGAGTG 354 QY 308 AGCCCAAGATTGCACCACTGCATCCAGCTCCAGCCCTGGAAAAAAGGGTAA 352 Db 355 AGCTAAGATCGTGCCATTGCACTTCAGCTCCAGCCCTGGACAACACAGAGTGA 399</p>		<p>RESULT 15 BE169870/c LOCUS CM3-HT0528-010200-086-f04 HT0528 Homo sapiens cDNA, mRNA sequence. BE169870 DEFINITION BE169870.1 GI:8632591 ACCESSION BE169870.1 KEYWORDS EST. SOURCE Homo sapiens (human)</p>	
<p>TITLE 21C Frontier Korean EST Project 2001 JOURNAL Unpublished (2002) COMMENT Contact: Kim YS Genome Research Center Korea Research Institute of Bioscience & Biotechnology 52 Boeun-dong Yuseong-gu, Daejeon 305-333, South Korea Tel: +82-42-860-4470 Fax: +82-42-860-4409 Email: yongsung@mail.kribb.re.kr Plate: 20 row: H column: 06 High quality sequence stop: 657.</p>		<p>FEATURES Location/Qualifiers source 1..657 /organism="Homo sapiens" /mol_type="mRNA" /db_xref="taxon:9606" /clone="S10SNU1-20-H06" /sex="M" /tissue_type="Stomach" /cell_type="Lymphoblast-like" /cell_line="SNU-1" /lab_host="Top10P" /clone_lib="S10SNU1" /note="Organ: Stomach; Vector: pME18-FL3; Site 1: XhoI; Site 2: XhoI; The poly (A)+ RNA was dephosphorylated with bacterial alkaline phosphatase (BAP) and then decapped with tabacco acid pyrophosphatase (TAP). The decapped intact mRNA was ligated with DNA-RNA linker including SfiI site by treatment of T4 RNA ligase and the first strand cDNA was synthesized with Superscript II using SfiI oligo-dT primer. After first strand synthesis, RNA was degraded by NaOH treatment and cDNA was amplified by PCR reaction. The PCR products were digested with SfiI and cloned into DraIII- digested pME18S-FL3 vector. The obtained cDNA vectors were used for transformation of competent cells E. coli Top10P by electroporation method. The cDNA libraries constructed by this method are full-length enriched cDNA library."</p>		<p>ORIGIN Query Match 41.4%; Score 206.6; DB 3; Length 657; Best Local Similarity 87.0%; Pred. No. 3.1e-25; Matches 227; Conservative 0; Mismatches 34; Indels 0; Gaps 0; QY 79 TTAGCCAGGTGGGTGGCTCATGCTTATTTCCAGCACTTTGGGAGACCGAGTTGGGTG 138 Db 425 TCAGCCAGGCGGAGTGGCTCAGCCTGTGTAATCCAGCACTTTGGGAGGCGGAGTGGGTG 366 QY 139 GATCACTTGAGAGCAGGAGTTTGGAGACCACTGGCCCAACATGGTGAACCCCATCTCTA 198 Db 365 GATCACTTGAGGTGAGAGTTTGGAGACCACTGGCCCAACATGGTGAACCCCATCTCTA 306 QY 199 ATAAAAATCAAAAAATTAGCAGGTGTGGTGGCACTTGCCTATAGTCCCAGCTACACAGG 258 Db 305 CTAAAAATCAAAAAATTAGCAGGATGTGGTGGCATGTGCTCTTAACCCAGCTACTTGGG 246 QY 259 AGGCTGGGCGCAGAGAAGATCACTTGAACCTGGAGGTGGAGGTGGAGTGGAGTGGAGTGG 318 Db 245 AGGCTGAGGCGAGGAAATCACTTGAACCCAGGAGTGGAGGTGGAGTGGAGTGGAGTGG 186 QY 319 CACCACTGCACTCCAGCTGG 339 Db 185 TGCCACTGTACTCCAGCCTGG 165</p>		<p>RESULT 14 CR960649 LOCUS CR960649 795 bp DNA linear GSS 06-JUN-2005 DEFINITION Homo sapiens BAC end sequence of RZPDB737B062021D from genomic library (orig. Pieter J. de Jong library RZPD-11), genomic survey sequence. ACCESSION CR960649 VERSION CR960649.1 GI:66991211</p>					

THIS PAGE BLANK (USPTO)

GenCore version 5.1.7

Copyright (c) 1993 - 2006 Bioceleration Ltd.

OM nucleic - nucleic search, using sw model

Run on: February 9, 2006, 13:28:05 ; Search time 2697.09 Seconds
(without alignments)
10559.003 Million cell updates/sec

Title: US-10-607-806-1_COPY_5000_5500

Perfect score: 501

Sequence: 1 acggggctcaactctgtgcc.....ttaataaagacgggggtttc 501

Scoring table: IDENTITY NUC

Gapop 10_0 , Gapext 1.0

Searched: 5893141 seqs, 28421725653 residues

Total number of hits satisfying chosen parameters: 11766282

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

GenEmbl.*

1: gb_ba.*

2: gb_in.*

3: gb_env.*

4: gb_om.*

5: gb_ov.*

6: gb_pat.*

7: gb_ph.*

8: gb_pr.*

9: gb_ro.*

10: gb_sts.*

11: gb_sy.*

12: gb_un.*

13: gb_vl.*

14: gb_htg.*

15: gb_pl.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	501	100.0	8368	8 AY438977	AY438977 Homo sapi
2	501	100.0	13612	6 AX377239	AX377239 Sequence
C 3	501	100.0	122302	8 AC003982	AC003982 Homo sapi
C 4	501	100.0	220384	14 AC078926	AC078926 Homo sapi
5	429.2	85.7	3375	8 HUMPLA2A1	M22970 Human pancr
C 6	248.8	49.7	112069	8 AL356103	AL356103 Human DNA
7	239.8	47.9	153881	14 AC011793	AC011793 Homo sapi
C 8	239.8	47.9	181655	14 AC024488	AC024488 Homo sapi
9	238.6	47.6	193352	14 AC116046	AC116046 Papio ham
10	237	47.3	123253	8 AC004971	AC004971 Homo sapi
C 11	236.8	47.3	158905	8 AC005666	AC005666 Homo sapi
C 12	236.8	47.3	170892	14 AC068438	AC068438 Homo sapi
C 13	235.2	46.9	174902	8 AC099811	AC099811 Homo sapi
14	234.8	46.9	107529	8 AC1357936	AC1357936 Human DNA
15	234.2	46.7	150151	8 AL356115	AL356115 Human DNA
C 16	234.2	46.7	211001	8 AC018695	AC018695 Homo sapi
C 17	233.8	46.7	203134	8 AC021077	AC021077 Homo sapi
18	233.6	46.6	163331	8 AC092802	AC092802 Homo sapi

19	233.6	46.6	171641	8 AC018512	AC018512 Homo sapi
C 20	233.6	46.6	172421	14 AC013279	AC013279 Homo sapi
21	233.6	46.6	196603	8 AC087721	AC087721 Homo sapi
22	233.4	46.6	62613	14 AC025789	AC025789 Homo sapi
C 23	233.4	46.6	125625	8 DQ015919	DQ015919 Homo sapi
C 24	233.4	46.6	143605	8 AC005522	AC005522 Homo sapi
25	233.4	46.6	188117	8 AC022613	AC022613 Homo sapi
C 26	233.4	46.6	201215	8 AC019084	AC019084 Homo sapi
27	233.4	46.4	80586	8 AC012574	AC012574 Homo sapi
C 28	231.8	46.3	135996	14 AL356007	AL356007 Homo sapi
29	231.6	46.2	124645	8 HUAC004638	AC004638 Homo sapi
30	231.6	46.2	124998	8 AC026397	AC026397 Homo sapi
C 31	231.6	46.2	125099	8 AC006251	AC006251 Homo sapi
32	231.6	46.2	128877	14 AC026399	AC026399 Homo sapi
33	231.6	46.2	170691	8 AC044802	AC044802 Homo sapi
C 34	231.4	46.2	171713	14 AL356152	AL356152 Homo sapi
C 35	230.8	46.1	164125	8 AC018693	AC018693 Homo sapi
C 36	230.6	46.0	179390	14 AC115532	AC115532 Papio anu
C 37	230.4	46.0	92518	8 BX088650	BX088650 Human DNA
C 38	230.2	45.9	193766	8 AC093420	AC093420 Homo sapi
39	230.2	45.9	200881	14 AC116566	AC116566 Homo sapi
40	230	45.9	206102	8 AC027131	AC027131 Homo sapi
41	229.8	45.9	184762	14 AC023500	AC023500 Homo sapi
C 42	229.6	45.8	163183	8 AC016574	AC016574 Homo sapi
C 43	229.6	45.8	187691	8 AL139082	AL139082 Human DNA
C 44	229.2	45.7	143506	8 AC092833	AC092833 Homo sapi
C 45	229	45.7	167309	14 AC157857	AC157857 Papio anu

ALIGNMENTS

RESULT 1	AY438977	8368 bp	DNA	linear	PRI 29-OCT-2003
LOCUS	AY438977	Homo sapiens phospholipase A2, group IB (pancreas)			(PLA2G1B) Gene,
DEFINITION	complete cds.				
ACCESSION	AY438977				
VERSION	AY438977.1	GI:37953284			
KEYWORDS					
SOURCE	Homo sapiens (human)				
ORGANISM	Homo sapiens				
REFERENCE	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini; Hominidae; Homo.				
AUTHORS	Rieder, M.J., Livingston, R.J., Daniels, M.R., Chung, M.-W., Miyamoto, K.E., Nguyen, C.P., Nguyen, D.A., Poel, C.L., Robertson, P.D., Schackwitz, W.S., Sherwood, J.K., Witrak, L.A. and Nickerson, D.A.				
TITLE	Submitted (16-OCT-2003) Genome Sciences, University of Washington, 1705 NE Pacific, Seattle, WA 98195, USA				
JOURNAL	Project, NIEHS ES15478, Department of Genome Sciences, Seattle, WA (URL: http://egp.gs.washington.edu).				
COMMENT	Location/Qualifiers				
FEATURES	source	1..8368			
	organism="Homo sapiens"				
	/mol_type="genomic DNA"				
	/db_xref="taxon:9606"				
	101..257				
repeat_region	/rpt_family="MIR"				
	/rpt_type="dispersed"				
variation	205				
	/frequency="0.01"				
	/replace="a"				
variation	306				
	/frequency="0.01"				
	/replace="a"				
repeat_region	326..538				
	/rpt_family="Alu"				
	/rpt_type="dispersed"				
misc_feature	354..496				

gene	variation
875. .6548	/note="Region not scanned for variation"
/gene="PLA2G1B"	/gene="PLA2G1B"
join(875. .944,2644. .2803,3603. .3730,6347. .6548)	/frequency="0.10"
/gene="PLA2G1B"	/replace="c"
3702	variation
/gene="PLA2G1B"	/gene="PLA2G1B"
product="phospholipase A2, group IB (pancreas)"	/frequency="0.15"
join(911. .944,2644. .2803,3603. .3730,6347. .6471)	/replace="a"
/gene="PLA2G1B"	variation
/codon_start=1	3768
/product="phospholipase A2, group IB (pancreas)"	/gene="PLA2G1B"
/protein_id="AA05441.1"	/frequency="0.01"
/db_xref="GI:37953285"	/replace="t"
3818	variation
/translation="MKLLVLAIVLTVAADSGISPRAWQFRKMIKCVIPGSDPFLFY	/gene="PLA2G1B"
NNYCYGLSGSGTPVDLDRCCQTHDNCYDQAKLDCKFLLDNPYTHITSYSCSGS	/frequency="0.01"
AITCSSKKECEAFICNDRNAATCFSKAPYNAKKNLDTKYQCS"	/replace="a"
1174. .1307	repeat_region
/rpt_family="Alu"	3846. .3964
/rpt_type="dispersed	/rpt_family="MIR"
1286	/rpt_type="dispersed
/gene="PLA2G1B"	3977. .4280
/frequency="0.01"	/rpt_family="Alu"
/replace="g"	/rpt_type="dispersed
1319. .1588	repeat_region
/rpt_family="Alu"	4281. .4346
/rpt_type="dispersed	/rpt_family="L2"
1343	/rpt_type="dispersed
/gene="PLA2G1B"	4347. .4634
/frequency="0.03"	/rpt_family="Alu"
/replace="a"	/rpt_type="dispersed
1593	variation
/gene="PLA2G1B"	4386
/frequency="0.02"	/gene="PLA2G1B"
/replace="t"	/frequency="0.08"
1605. .1669	/replace="t"
/rpt_family="L2"	4635. .4723
/rpt_type="dispersed	/rpt_family="L2"
1638. .2188	/rpt_type="dispersed
/gene="PLA2G1B"	5013
/note="Region not scanned for variation"	/gene="PLA2G1B"
1670. .1975	/frequency="0.10"
/rpt_family="Alu"	/replace="a"
/rpt_type="dispersed	5103. .5404
1976. .2038	/rpt_family="Alu"
/rpt_type="dispersed	/rpt_type="dispersed
2086. .2165	5133
/rpt_family="L2"	/gene="PLA2G1B"
/rpt_type="dispersed	/frequency="0.01"
2194. .2319	/replace="c"
/rpt_family="Alu"	5543
/rpt_type="dispersed	/gene="PLA2G1B"
2319. .2453	/frequency="0.01"
/gene="PLA2G1B"	/replace="t"
/note="Region not scanned for variation"	5556
3027	/gene="PLA2G1B"
/frequency="0.25"	/frequency="0.15"
/replace="t"	/replace="g"
3165. .3431	5577. .5629
/rpt_family="Alu"	/rpt_family="L1"
/rpt_type="dispersed	/rpt_type="dispersed
3262	5580
/gene="PLA2G1B"	/gene="PLA2G1B"
/frequency="0.01"	/frequency="0.09"
/replace="t"	/replace="g"
3321	5675
/gene="PLA2G1B"	/gene="PLA2G1B"
/frequency="0.01"	/frequency="0.09"
/replace="a"	/replace="g"
3396	5779. .6083
/gene="PLA2G1B"	/rpt_family="Alu"
/frequency="0.03"	/rpt_type="dispersed
/replace="g"	5929

/gene="PLA2G1B"
/frequency="0.01"
/replace="a"
5949
/gene="PLA2G1B"
/frequency="0.01"
/replace="g"
6039
/gene="PLA2G1B"
/frequency="0.01"
/replace="a"
6082
/gene="PLA2G1B"
/frequency="0.14"
/replace="t"

Query Match 100.0%; Score 501; DB 8; Length 8368;
Best Local Similarity 100.0%; Pred. No. 3.5e-131; Indels 0; Gaps 0;
Matches 501; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 ACGGGGCTCACTCTGTGCGCCAGGCTGGAGTGGCGGTGATCTCAGATCACTACAC 60
DB 1374 ACGGGGCTCACTCTGTGCGCCAGGCTGGAGTGGCGGTGATCTCAGATCACTACAC 1433
QY 61 CTCATCTCTGGGTTCAAAATAATTCCTTGCTCAGCTCTCAAGTAGCTGGGACTTGT 120
DB 1434 CTCATCTCTGGGTTCAAAATAATTCCTTGCTCAGCTCTCAAGTAGCTGGGACTTGT 1493
QY 121 AGCTCTCAAGTAGCTGGGTTCAAAATAATTCCTTGCTCAGCTCTCAAGTAGCTGGGACTTGT 180
DB 1494 AGCTCTCAAGTAGCTGGGTTCAAAATAATTCCTTGCTCAGCTCTCAAGTAGCTGGGACTTGT 1553
QY 181 GTAGACAGAGTTTTCACCATGTTGGCCAGGCTGGGTGACCTCCCTTTTAGATTCTCCTC 240
DB 1554 GTAGACAGAGTTTTCACCATGTTGGCCAGGCTGGGTGACCTCCCTTTTAGATTCTCCTC 1613
QY 241 ATCTGCTCTATCTTCCCTTTTCAATGCGAGTATCCAGTTTCTTACTTATCAATTA 300
DB 1614 ATCTGCTCTATCTTCCCTTTTCAATGCGAGTATCCAGTTTCTTACTTATCAATTA 1673
QY 301 TTATTAATTTTATTAATTTAGACAGAGTCTTGTGTTGTCGCCAAGGCTGGAGTACAGT 360
DB 1674 TTATTAATTTTATTAATTTAGACAGAGTCTTGTGTTGTCGCCAAGGCTGGAGTACAGT 1733
QY 361 GTGCGATCTGGCTCACTGCAAGCTCCACCTGCTGGGTTTCAAGGCTTCTCCGCTCA 420
DB 1734 GTGCGATCTGGCTCACTGCAAGCTCCACCTGCTGGGTTTCAAGGCTTCTCCGCTCA 1793
QY 421 GCCTCCCGAGTAGCTGGGACTAAAGGCGCTGCCACCGCCGCTAAATTTTGTAT 480
DB 1794 GCCTCCCGAGTAGCTGGGACTAAAGGCGCTGCCACCGCCGCTAAATTTTGTAT 1853
QY 481 TTTTAATAAAGACGGGGTTTC 501
DB 1854 TTTTAATAAAGACGGGGTTTC 1874

RESULT 2
AX377239
LOCUS AX377239 13612 bp DNA linear PAT 18-MAR-2002
DEFINITION Sequence 1 from Patent WO212562.
ACCESSION AX377239
VERSION AX377239.1 GI:19573528
KEYWORDS Homo sapiens (human)
SOURCE Homo sapiens
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini;
Hominidae; Homo.
1
REFERENCE
AUTHORS Kazemi, A., Kliehm, S.E. and Koshiy, B.
TITLE Haplotypes of the pla2g1b gene
JOURNAL Patent: WO 0212562-A 1 14-FEB-2002;

Genaisance Pharmaceuticals, Inc. (US)
Location/Qualifiers
1. 13612
/organism="Homo sapiens"
/mol_type="unassigned DNA"
/db_xref="taxon:9606"
3845
/note="PS1: polymorphic base G or A"
3968
/note="PS2: polymorphic base A or G"
6060
/note="PS2: polymorphic base A or G"
6844
/note="PS4: polymorphic base G or A"
9531
/note="PS5: polymorphic base G or A"
ORIGIN
Query Match 100.0%; Score 501; DB 6; Length 13612;
Best Local Similarity 100.0%; Pred. No. 3.5e-131;
Matches 501; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 ACGGGGCTCACTCTGTGCGCCAGGCTGGAGTGGCGGTGATCTCAGATCACTACAC 60
DB 4516 ACGGGGCTCACTCTGTGCGCCAGGCTGGAGTGGCGGTGATCTCAGATCACTACAC 4575
QY 61 CTCATCTCTGGGTTCAAAATAATTCCTTGCTCAGCTCTCAAGTAGCTGGGACTTGT 120
DB 4576 CTCATCTCTGGGTTCAAAATAATTCCTTGCTCAGCTCTCAAGTAGCTGGGACTTGT 4635
QY 121 AGCTCTCAAGTAGCTGGGTTCAAAATAATTCCTTGCTCAGCTCTCAAGTAGCTGGGACTTGT 180
DB 4636 AGCTCTCAAGTAGCTGGGTTCAAAATAATTCCTTGCTCAGCTCTCAAGTAGCTGGGACTTGT 4695
QY 181 GTAGACAGAGTTTTCACCATGTTGGCCAGGCTGGGTGACCTCCCTTTTAGATTCTCCTC 240
DB 4696 GTAGACAGAGTTTTCACCATGTTGGCCAGGCTGGGTGACCTCCCTTTTAGATTCTCCTC 4755
QY 241 ATCTGCTCTATCTTCCCTTTTCAATGCGAGTATCCAGTTTCTTACTTATCAATTA 300
DB 4756 ATCTGCTCTATCTTCCCTTTTCAATGCGAGTATCCAGTTTCTTACTTATCAATTA 4815
QY 301 TTATTAATTTTATTAATTTAGACAGAGTCTTGTGTTGTCGCCAAGGCTGGAGTACAGT 360
DB 4816 TTATTAATTTTATTAATTTAGACAGAGTCTTGTGTTGTCGCCAAGGCTGGAGTACAGT 4875
QY 361 GTGCGATCTGGCTCACTGCAAGCTCCACCTGCTGGGTTTCAAGGCTTCTCCGCTCA 420
DB 4876 GTGCGATCTGGCTCACTGCAAGCTCCACCTGCTGGGTTTCAAGGCTTCTCCGCTCA 4935
QY 421 GCCTCCCGAGTAGCTGGGACTAAAGGCGCTGCCACCGCCGCTAAATTTTGTAT 480
DB 4936 GCCTCCCGAGTAGCTGGGACTAAAGGCGCTGCCACCGCCGCTAAATTTTGTAT 4995
QY 481 TTTTAATAAAGACGGGGTTTC 501
DB 4996 TTTTAATAAAGACGGGGTTTC 5016

RESULT 3
AC003982/c
LOCUS AC003982 122302 bp DNA linear PRI 18-MAR-1999
DEFINITION Homo sapiens PAC clone 166H1 from 12q, complete sequence.
ACCESSION AC003982
VERSION AC003982.1 GI:2769695
KEYWORDS HTG.
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini;
Hominidae; Homo.
1 (bases 1 to 122302)
REFERENCE
AUTHORS Sulston, J.E. and Waterston, R.


```

TITLE
JOURNAL      Toward a complete human genome sequence
PUBMED       Genome Res. 8 (11), 1097-1108 (1998)
REFERENCE    9847074
2 (bases 1 to 122302)
Bradshaw,H., Wu,X. and Ozersky,P.
TITLE        The sequence of Homo sapiens PAC clone 166H1
JOURNAL      Unpublished (1999)
REFERENCE    3 (bases 1 to 122302)
Waterston,R.
Direct Submission
Submitted (13-JAN-1998) Department of Genetics, Washington
University, 4444 Forest Park Avenue, St. Louis, Missouri 63108, USA
4 (bases 1 to 122302)
Waterston,R.
Direct Submission
Submitted (18-MAR-1999) Department of Genetics, Washington
University, 4444 Forest Park Avenue, St. Louis, Missouri 63108, USA
SUBMITTED BY: WUGSC
Genome Sequencing Center
Department of Genetics
Washington University
St. Louis MO 63108, USA
http://genome.wustl.edu/gsc
mailto:sapiens@watson.wustl.edu

NOTICE: This sequence may not represent the entire insert of this
clone. It may be shorter because we only sequence overlapping
clone sections once, or longer because we provide a small overlap
between neighboring data submissions.

This sequence was finished as follows unless otherwise noted:
all regions were double stranded, sequenced with an alternate
chemistry, or covered by high quality data (i.e., phred quality >=
30); an attempt was made to resolve all sequencing problems, such
as compressions and repeats; all regions were covered by sequence
from more than one subclone; and the assembly was confirmed by
restriction digest.

MAPPING INFORMATION:
This clone was originally isolated in the laboratory of Professor
Graeme Bell, Howard Hughes Medical Institute and Departments of
Biochemistry and Molecular Biology, and Medicine, The University of
Chicago, Chicago, IL, USA. The clone was provided by the
laboratory of Dr. Roger Cox at the Wellcome Trust Centre For Human
Genetics, Oxford, UK. Some contig information was also obtained
from Yamagata et al., Nature 384:455-8 (1996).

SOURCE INFORMATION:
This clone was derived from human PAC library RPCI-1, prepared by
Pieter de Jong and coworkers at the Roswell Park Cancer Institute
(http://bacpac.med.buffalo.edu) using the method described by
Ioannou et al., Nature Genetics 6:84-9 (1994). The library is from
one male donor.
The clone may be obtained either from Genome Systems, Inc.
(http://www.genomesystems.com) or Research Genetics, Inc.
(http://www.resgen.com); or from Pieter de Jong.
VECTOR: pCYPAC2

NEIGHBORING SEQUENCE INFORMATION:
The clone sequenced to the left is 278C19; the clone sequenced to
the right is 15E1. Actual start of this clone is at base position
1 of 166H1; actual end is at 122302 of 166H1.
1 of 166H1; actual end is at 122302 of 166H1.
Location/Qualifiers
1. .122302
/organism="Homo sapiens"
/mol_type="genomic DNA"
/db_xref="taxon:9606"
/chromosome="12"
/map="12q"
/clone="166H1"
/clone_lib="RPCI-1"
repeat_region 1. .238
/rpt_family="Alu"
repeat_region 239. .265

```

```

/rpt_family="AT_rich"
268. .565
/rpt_family="Alu"
1001. .1309
/rpt_family="Alu"
1440. .1745
/rpt_family="Alu"
1909. .1940
/rpt_family="MERI_type"
1955. .1983
/rpt_family="AT_rich"
2063. .2362
/rpt_family="Alu"
2516. .2650
/rpt_family="Alu"
2674. .2730
/rpt_family="L1"
2735. .2788
/rpt_family=" (GAAAA)n"
2790. .3077
/rpt_family="Alu"
3081. .3139
/rpt_family="L1"
3716. .4020
/rpt_family="Alu"
4196. .4492
/rpt_family="Alu"
4560. .14034
/gene="WUGSC:H.166H1.1"
join(4560. .5056,13418. .13748,13882. .14034)
/gene="WUGSC:H.166H1.1"
/note="unknown function; 60% similar to Z50177
(PID:G927403) (PID:G927402); H_166H1.1"
/codon_start=1
/evidence=not experimental
/protein_id="AAB95634.1"
/db_xref="GI:2769696"
/translation="MKMSFALTPRSAGRWIANSPQCSKASIGLVFVSPPLDPEKV
KELORPFTLSKRLVMTGAGISTESGIPDYSEKVGLYARTDRPIQHGDFVRSAPIR
QRYWARNFVGFQSSSQPNPAHWALSTWEKLGKLYWLVTQNDALHTKAGSRRLTEL
HGMCDRAYCSVSFVLSGVLCDCEQTPRGVLQRFQVLTNPWTSAEAGLAPDGDVF
LSEQVRSQVFTVCQCGHLKPDVTFGDTVPDKVDVHKRVKREASDLLVVGSSLQ
VYSGVRFILTAKKKLPAILNIGPTRSDDLACLKLNRCGELLPLIDPC"
5366. .5655
/rpt_family="Alu"
5672. .5798
/rpt_family="Alu"
5802. .5834
/rpt_family=" (TA)n"
5834. .5857
/rpt_family="POLY_A"
5870. .6184
/rpt_family="Alu"
6224. .6268
/rpt_family=" (CA)n"
6269. .6558
/rpt_family="Alu"
6729. .7029
/rpt_family="Alu"
7051. .7172
/rpt_family="Alu"
7192. .7500
/rpt_family="Alu"
7518. .7820
/rpt_family="Alu"
8207. .8500
/rpt_family="Alu"
8576. .8878
/rpt_family="Alu"
9023. .9154
/rpt_family="Alu"
9165. .9470
/rpt_family="Alu"

```

```

repeat_region 9887..10188
/rpt_family="Alu"
repeat_region 10221..10342
/rpt_family="L1"
repeat_region 10408..10707
/rpt_family="Alu"
repeat_region 10713..11014
/rpt_family="Alu"
repeat_region 11020..11110
/rpt_family="Alu"
repeat_region 11146..11386
/rpt_family="Alu"
repeat_region 11208..11386
/rpt_family="Alu"
repeat_region 11391..11515
/rpt_family="Alu"
repeat_region 11568..11705
/rpt_family="Alu"
repeat_region 11737..12032
/rpt_family="Alu"
repeat_region 12034..12066
/rpt_family="7SLRNA"
repeat_region 12075..12372
/rpt_family="Alu"
repeat_region 12448..12753
/rpt_family="Alu"
repeat_region 12763..12874
/rpt_family="L2"
repeat_region 12999..13303
/rpt_family="Alu"
misc_feature complement(13674..13752)
/note="match to EST N80020 (NID:g1242721) za91a08.s1"
misc_feature complement(13840..14195)
/note="match to EST N80020 (NID:g1242721) za91a08.s1"
repeat_region 14196..14229
/rpt_family="AT-rich"
repeat_region 14240..14339
/rpt_family="U6"
repeat_region 14349..14646
/rpt_family="Alu"
repeat_region 14652..14940
/rpt_family="Alu"

Query Match 100.0%; Score 501; DB 8; Length 122302;
Best Local Similarity 100.0%; Pred. No. 3.7e-131;
Matches 501; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 ACGGGGCTCATCTGTGTCGCCAGGCTGGAGTGGCGTGGATCTCAGATCACTACAAC 60
Db 28288 ACGGGGCTCATCTGTGTCGCCAGGCTGGAGTGGCGTGGATCTCAGATCACTACAAC 28229

QY 61 CTCATCTCTCGGGTCAAAATAATCTCTGCTCAGCTCTCAAGTAGCTGGACTGT 120
Db 28228 CTCATCTCTCGGGTCAAAATAATCTCTGCTCAGCTCTCAAGTAGCTGGACTGT 28169

QY 121 AGCTCTCAAGTAGCTGGACACACACACACATCCAGCTAAATTTTGTGTTTTTTT 180
Db 28168 AGCTCTCAAGTAGCTGGACACACACACATCCAGCTAAATTTTGTGTTTTTTT 28109

QY 181 GTAGACAGGTTTTTACCATTGTTGGCAGGCTGGTGAACCTCCTTTTGTAGATTCCTC 240
Db 28108 GTAGACAGGTTTTTACCATTGTTGGCAGGCTGGTGAACCTCCTTTTGTAGATTCCTC 28049

QY 241 ATCTGCTCTATTTCTCCCTTTCTAATGAGTATCCAGTTCTCTTACTATCACTTTA 300
Db 28048 ATCTGCTCTATTTCTCCCTTTCTAATGAGTATCCAGTTCTCTTACTATCACTTTA 27989

QY 301 TTATTATCTTATTATTATTGAGACAGTCTTCTTTGTGCGCAAGCTGGAGTACAGT 360
Db 27988 TTATTATCTTATTATTATTGAGACAGTCTTCTTTGTGCGCAAGCTGGAGTACAGT 27929

QY 361 GGTGCGATCTGGCTCACTCAAGCTCCACTGCTGGGTTTCAAGCCATTCTCCGCGCTCA 420
|||||

```

```

Db 27928 GGTGGGATCTCGGCTCACTGCAAGCTCCACCTGCTGGGTTCACGCACTCTCCGCTCA 27869
QY 421 GCCTCCCCAGTAGCTGGGACTAAAGGCGCTGCCACACGCGCCGCTAATTTTGTAT 480
|||||
Db 27868 GCCTCCCCAGTAGCTGGGACTAAAGGCGCTGCCACACGCGCCGCTAATTTTGTAT 27809
QY 481 TTTTAATAAAGACGGGGTTTC 501
|||||
Db 27808 TTTTAATAAAGACGGGGTTTC 27788

```

```

RESULT 4
AC078926/c
LOCUS
DEFINITION Homo sapiens chromosome 12 clone RP11-836M11, WORKING DRAFT
SEQUENCE, 11 unordered pieces.
AC078926
AC078926.23 GI:19718703
VERSION HTG; HTGS PHASE1; HTGS_DRAFT; HTGS_FULLTOP.
KEYWORDS Homo sapiens (human)
SOURCE
ORGANISM Homo sapiens

```

```

Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini;
Hominoidea; Homo.
1 (bases 1 to 220384)
AC078926
Muzny,D.M., Adams,C., Adio-Oduola,B., Ali-osman,F.R., Allen,C.,
Alsbrooks,S.L., Amaratunge,H.C., Are,J.R., Ayele,M., Banks,T.,
Barbaria,J., Benton,J., Bimage,K., Blankenburg,K., Bonnin,D.,
Bouck,J., Bowie,S., Brieva,M., Brown,E., Brown,M., Bryant,N.P.,
Buhay,T., Burch,P., Burkett,C., Burrell,K.B., Byrd,N.C.,
Carron,T.F., Carter,M., Cavazos,S.R., Chacko,J., Chavez,D.,
Chen,G., Chen,R., Chen,Z., Chowdhry,I., Christopoulos,C.,
Cleveland,C.D., Cox,C., Coyle,M.D., Dathorne,S.R., David,R.,
Delava,M.L., Davis,C., Davy-Carroll,L., Dederich,D.A.,
Delaney,K.R., Delgado,O., Denn,A.L., Ding,Y., Dinh,H.H.,
Douthwaite,K.J., Draper,H., Dugan-Rocha,S., Durbin,K.J.,
Earnhart,C., Edgar,D., Edwards,C.C., Elhaj,C., Escotto,M.,
Falls,T., Ferraguto,D., Flagg,N., Ford,J., Foster,P., Frantz,P.,
Gabisi,A., Gao,J., Garcia,A., Garner,T., Garza,N., Gill,R.R.,
Gorrell,J.H., Guevara,M., Gunaratne,P., Hale,S., Hamilton,K.,
Harris,C., Harris,K., Hart,M., Havlak,P., Hawes,A., Hernandez,J.,
Hernandez,O., Hodgson,A., Hogue,M., Holloway,C., Hollins,B.,
Homs,F., Howard,S., Huber,J., Hulyk,S., Hume,J., Jackson,L.E.,
Jacomson,B., Jia,Y., Johnson,R., Jolivet,S., Joudah,S.,
Karleson,E., Kelly,S., Khan,U., King,L., Korvah,J., Kovar,C.,
Kratovic,J., Kureshi,A., Landry,N., Leal,B., Lewis,L.C., Lewis,L.,
Lozado,R.J., Lu,X., Lucier,A., Lucier,R., Luna,R., Ma,J.,
Maheshwari,M., Mapua,P., Martin,R., Martindale,A., Martinez,E.,
Massey,E., Mawhney,E., McLeod,M.P., Meador,M., Mei,G., Metzker,M.,
Miner,G., Miner,Z., Mitchell,T., Mohabbat,K., Morgan,M., Morris,S.,
Moser,M., Neal,D., Newton,J., Newton,N., Nguyen,A., Nguyen,N.,
Nguyen,N., Nickerson,B., Nwokenwo,S., Ogih,M., Okwuonu,G.,
Oguyunye,N., Oviedo,R., Pace,A., Payton,B., Peery,J., Perez,L.,
Peters,L., Pickens,R., Primus,E., Pu,L.L., Quiles,M., Ren,Y.,
Rives,M., Rojas,A., Rojubokan,I., Rolfe,M., Ruiz,S., Savery,G.,
Scherer,S., Scott,G., Shen,H., Shoohetari,N., Sisson,I.,
Suttongren,E., Sonaika,T., Sparks,A., Stanley,H., Stone,H.,
Suttongren,A., Svatek,A., Taber,P., Tamerisa,A., Tamerisa,K., Tang,H.,
Tansey,J., Taylor,C., Taylor,T., Telford,B., Thomas,N., Thomas,S.,
Umani,K., Vasquez,L., Vera,V., Villalob,D., Vinson,R., Wang,Q.,
Wang,S., Ward-Moore,S., Warren,R., Washington,C., Wallington,S.,
Williams,G., Williamson,A., Wlezyk,R., Wooden,S., Worley,K.,
Wu,C., Wu,Y., Wu,Y.F., Zhou,J., Zorrilla,S., Nelson,D.,
Weinstock,G., and Gibbs,R.
Direct Submission
Unpublished
2 (bases 1 to 220384)
Worley,K.C.
Direct Submission
Submitted (11-AUG-2000) Human Genome Sequencing Center, Department
of Molecular and Human Genetics, Baylor College of Medicine, One
Baylor Plaza, Houston, TX 77030, USA

```

```

TITLE
JOURNAL
REFERENCE
AUTHORS
TITLE
JOURNAL

```

REFERENCE 3 (bases 1 to 220384)
AUTHORS Worley,K.C.
TITLE Direct Submission
JOURNAL Submitted (26-MAR-2002) Human Genome Sequencing Center, Department of Molecular and Human Genetics, Baylor College of Medicine, One Baylor Plaza, Houston, TX 77030, USA
COMMENT On Mar 26, 2002 this sequence version replaced gi:18449664.
----- Genome Center
Center: Baylor College of Medicine
Center code: BCM
Web site: http://www.hgsc.bcm.tmc.edu/
Contact: hgsc-help@bcm.tmc.edu
----- Project Information
Center project name: HCM
Center clone name: RP11-836M11
----- Summary Statistics
Sequencing vector: Plasmid;
Sequencing vector: M13;
Chemistry: Dye-primer Bodipy: 47% of reads
Chemistry: Dye-terminator Big Dye: 53% of reads
Assembly program: Phrap; version 0.990329
Consensus quality: 234783 bases at least Q40
Consensus quality: 241821 bases at least Q30
Consensus quality: 245671 bases at least Q20
Estimated insert size: 219187; sum-of-contigs estimation
Quality coverage: 8.9x in Q20 bases; sum-of-contigs estimation

* NOTE: Estimated insert size may differ from sequence length
(see http://www.hgsc.bcm.tmc.edu/docs/Genbank_draft_data.html).
* NOTE: This is a 'working draft' sequence. It currently
* consists of 11 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.
*
* 1 2156: contig of 2156 bp in length
* 2157 2256: gap of unknown length
* 2257 7112: contig of 4856 bp in length
* 7113 7212: gap of unknown length
* 7213 13283: contig of 6071 bp in length
* 13284 13383: gap of unknown length
* 20660: contig of 7277 bp in length
* 20661 20760: gap of unknown length
* 20761 34473: contig of 13713 bp in length
* 34474 44394: gap of unknown length
* 44395 44494: gap of unknown length
* 44495 57126: contig of 12632 bp in length
* 57127 57226: gap of unknown length
* 57227 80266: contig of 23040 bp in length
* 80267 80366: gap of unknown length
* 80367 109398: contig of 29032 bp in length
* 109399 109499: gap of unknown length
* 109499 149812: contig of 40314 bp in length
* 149813 149913: gap of unknown length
* 149913 220384: contig of 70472 bp in length.
* Location/Qualifiers
1. 220384
/organism="Homo sapiens"
/mol_type="genomic DNA"
/db_xref="taxon:9606"
/chromosome="12"
/clone="RP11-836M11"
2157..2256
/estimated_length=unknown
7113..7212
/estimated_length=unknown
13284..13383
/estimated_length=unknown
20661..20760
/estimated_length=unknown
source
organism="Homo sapiens"
mol_type="genomic DNA"
db_xref="taxon:9606"
chromosome="12"
clone="RP11-836M11"
2157..2256
estimated_length=unknown
7113..7212
estimated_length=unknown
13284..13383
estimated_length=unknown
20661..20760
estimated_length=unknown

gap 34474..34573
/estimated_length=unknown
gap 44395..44494
/estimated_length=unknown
gap 57127..57226
/estimated_length=unknown
gap 80267..80366
/estimated_length=unknown
gap 109399..109498
/estimated_length=unknown
gap 149813..149912
/estimated_length=unknown
ORIGIN
Query Match 100.0%; Score 501; DB 14; Length 220384;
Best Local Similarity 100.0%; Pred. No. 3.7e-131; Indels 0; Gaps 0;
Matches 501; Conservative 0; Mismatches 0;
QY 1 ACGGGGCTCACTCTGTGCGCCAGGCTGGAGTGGGGTGGCGTGAATCTCAGATCACTACAAC 60
DB 73503 ACGGGGCTCACTCTGTGCGCCAGGCTGGAGTGGGGTGGCGTGAATCTCAGATCACTACAAC 73444
QY 61 CTCCATCTCCTGGGTTCAAATAAATCTCTTGGCCTCAGCCTCTCAAGTAGCTGGGACTTGT 120
DB 73443 CTCCATCTCCTGGGTTCAAATAAATCTCTTGGCCTCAGCCTCTCAAGTAGCTGGGACTTGT 73384
QY 121 AGCTCTCAAGTAGCTGGGACACACACACCATGCCAGCTAAATTTTGTGTTTTTTTGG 180
DB 73383 AGCTCTCAAGTAGCTGGGACACACACCATGCCAGCTAAATTTTGTGTTTTTTTGG 73324
QY 181 GTAGAGACAGGTTTTCACCATGTTGGCCAGGCTGGGTGACCTCCCTTTTGAATTCCTCTC 240
DB 73323 GTAGAGACAGGTTTTCACCATGTTGGCCAGGCTGGGTGACCTCCCTTTTGAATTCCTCTC 73264
QY 241 ATCTGCTCTATTCTTCCCTTTCTAATGACAGTATCCAGTTTCTTACTTATCATTTTA 300
DB 73263 ATCTGCTCTATTCTTCCCTTTCTAATGACAGTATCCAGTTTCTTACTTATCATTTTA 73204
QY 301 TTATTATTTCTATTATTATTATGACAGAGTCTTGTGTCGCAAGGCTGGAGTACAGT 360
DB 73203 TTATTATTTCTATTATTATTATGACAGAGTCTTGTGTCGCAAGGCTGGAGTACAGT 73144
QY 361 GGTGCGATCTCGGCTCACTGCAAGCTCCACCTGTGGGTTTCAAGCTTCTCCCGCTCA 420
DB 73143 GGTGCGATCTCGGCTCACTGCAAGCTCCACCTGTGGGTTTCAAGCTTCTCCCGCTCA 73084
QY 421 GCCTCCCGAGTAGCTGGGACTAAAGGGGCTTGGCCACACGCGCCGCTAAATTTTGTAT 480
DB 73083 GCCTCCCGAGTAGCTGGGACTAAAGGGGCTTGGCCACACGCGCCGCTAAATTTTGTAT 73024
QY 481 TTTTAATAAAGACGGGGTTTC 501
DB 73023 TTTTAATAAAGACGGGGTTTC 73003
RESULT 5
HUMPLA2A1
LOCUS Human pancreatic phospholipase A-2 (PLA-2) gene, exons 1 to 3.
DEFINITION M22970 M14965
ACCESSION M22970.1 GI:190008
VERSION phospholipase A2.
KEYWORDS 1 of 2
SEGMENT
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini;
Hominidae; Homo
1 (bases 1 to 3375)
Seilhamer,J.J., Randall,T.L., Yamanaka,M. and Johnson,L.K.
Pancreatic phospholipase A2: isolation of the human gene and cDNAs
from porcine pancreas and human lung
DNA 5 (6), 519-527 (1986)
REFERENCE
AUTHORS
TITLE
JOURNAL

PUBMED	3028739	source text: Homo sapiens pulverized liver DNA.
COMMENT	Original	Location/Qualifiers
FEATURES	source	1..3375
		/organism="Homo sapiens"
		/mol_type="genomic DNA"
		/db_xref="taxon:9606"
		/map="12q23-qter"
		/tissue_type="pulverized liver"
		<537..570
exon		/gene="PLA2A"
		/note="G00-120-715"
		/number=1
intron		571..2262
		/gene="PLA2A"
		/note="G00-120-715"
		/number=1
exon		2263..2422
		/gene="PLA2A"
		/note="G00-120-715"
		/number=2
intron		2423..3222
		/gene="PLA2A"
		/note="G00-120-715"
		/number=2
exon		3223..3350
		/gene="PLA2A"
		/note="G00-120-715"
		/number=3
intron		3351..3375
		/gene="PLA2A"
		/note="G00-120-715"
		/number=3
ORIGIN	5 bp upstream of PstI site.	
Query Match	85.7%;	Score 429.2; DB 8; Length 3375;
Best Local Similarity	97.4%;	Pred. No. 8.5e-111;
Matches	489; Conservative	0; Mismatches 8; Indels 5; Gaps 5;
QY	1	ACGGGGCTCCTCTGTCGCCAGGCTGGAGTCGGTGGCGTGCATCTCAGATCAGTACAC 60
Db	996	ACGGGGCTCCTCTGTCGCCAGGCTGGAGTCGGTGGCGTGCATCTCAGATCAGTACAC 1054
QY	61	CTCCATCTCTGGGTTCAATAATTTCTTGTGCTCAGCTCTCAAGTAGCTGGGACTTGT 120
Db	1055	CTCCATCTCTCTGGTCAATAATTTCTTGTGCTCAGCTCTCAAGTAGCTGGGACTTGT 1112
QY	121	AGCTCTCAAGTAGCTGGGACACACACATGCTCAATTTTGTGTTTGTGTTTGTGTTT 180
Db	1113	AGCTCTCAAGTAGCTGGGACACACACATGCTCAATTTTGTGTTTGTGTTTGTGTTT 1172
QY	181	GTAGAGACAGGTTTTCACATGTCGCCAGGCTGGTGCACCTCCCTTTAGATTTCTCCTC 240
Db	1173	GTAGAGACAGGTTTTCACATGTCGCCAGGCTGGTGCACCTCCCTTTAGATTTCTCCTC 1232
QY	241	ATCCTGCTTAFTTCTTCCCTTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCT 299
Db	1233	ATCCTGCTTAFTTCTTCCCTTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCT 1292
QY	300	ATT 359
Db	1293	ATT 1352
QY	360	TGGTCGATCTCGGCTCACTGCAAGCTCCAGTCTGGTGGTTCAGCCATTTCTCCGCTC 419
Db	1353	TGGTCGATCTCGGCTCACTGCAAGCTCCAGTCTGGTGGTTCAGCCATTTCTCCGCTC 1412
QY	420	AGCTTCCCGAGTAGCTGGGACTAAAGGGGCTTGCACACCGCCCGCTAAATTTTGTGTA 479
Db	1413	AGCTTCCCGAGTAGCTGGGACTAAAGGGGCTTGCACACCGCCCGCTAAATTTTGTGTA 1471
QY	480	TTTTTAATAAGACGGGGTTTC 501

Db 1472 TTTTAAATAAGACGGGGTTTC 1493

RESULT 6

AL356103 112069 bp DNA linear PRI 18-MAY-2005

Human DNA sequence from clone RP4-713A24 on chromosome 1p22.2-31.1

Contains STSS and GSSs, complete sequence.

AL356103

AL356103.8 GI:10178511

HTG.

Homo sapiens (human)

Homo sapiens

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini; Hominidae; Homo.

1 (bases 1 to 112069)

Corby,N.

Direct Submission

Submitted (13-MAY-2005) Wellcome Trust Sanger Institute, Hinxton, Cambridgeshire, CB10 1SA, UK. E-mail enquiries: vegas@sanger.ac.uk

Clone requests: clonerequest@sanger.ac.uk

On Sep 15, 2000 this sequence version replaced gi:9988355.

The following abbreviations are used to associate primary accession numbers given in the feature table with their source databases:

Em., EMBL; Sw., SWISSPROT; Tr., TRMBL; Wp., WORMPEP; Information on the WORMPEP database can be found at

http://www.sanger.ac.uk/Projects/C_elegans/wormpep

This sequence was generated from part of bacterial clone contigs of human chromosome 1, constructed by the Sanger Centre Chromosome 1 Mapping Group. Further information can be found at

http://www.sanger.ac.uk/HGP/Chr1

RP4-713A24 is from the library RPCI-4 constructed by the group of Pieter de Jong. For further details see

http://www.chori.org/bacpac/home.htm

VECTOR: pCYPAC2

----- Genome Center

Center: Wellcome Trust Sanger Institute

Center code: SC

Web site: http://www.sanger.ac.uk

Contact: vegas@sanger.ac.uk

This sequence was finished as follows unless otherwise noted: all regions were either double-stranded or sequenced with an alternate chemistry or covered by high quality data (i.e., phred quality >= 30); an attempt was made to resolve all sequencing problems, such as compressions and repeats; all regions were covered by at least one subclone; and the assembly was confirmed by restriction digest, except on the rare occasion of the clone being a YAC.

FEATURES

Location/Qualifiers

1..112069

/organism="Homo sapiens"

/mol_type="genomic DNA"

/db_xref="RZPD:RPCIF704A24713"

/db_xref="taxon:9606"

/chromosome="1"

/map="p22.2-31.1"

/clone="RP4-713A24"

/clone_lib="RPCI-4"

1

/note="Clone_left_end: RP4-713A24"

112069

/note="Clone_right_end: RP4-713A24"

misc_feature

misc_feature

ORIGIN

Query Match 49.7%; Score 248.8; DB 8; Length 112069;

Best Local Similarity 73.4%; Pred. No. 1.6e-59;

Matches 367; Conservative 0; Mismatches 112; Indels 21; Gaps 3;

QY 3 GGGGCTCCTCTGTGCGCCAGGCTGGAGTCGGTGGCGTGCATCTCAGATCAGTACACCT 62

Db 65537 GAGTCTCCTCTGTGTCGCCAGGCTGGAGTCAGTGGCATAAATTTGCGCTCAGTCAACCT 65478


```

misc_feature 92905..153881
              /note="assembly_fragment
              clone end:T7
              vector_side:right"
ORIGIN
Query Match      47.9%; Score 239.8; DB 14; Length 153881;
Best Local Similarity 72.3%; Pred. No. 5.8e-57;
Matches 362; Conservative 0; Mismatches 117; Indels 22; Gaps 3;
QY 3 GGGGCTCACTCTGTCGCCAGGCTGAGTGGCGTGGGATCTCAGATCAGTACCACT 62
Db 84323 GAGTCTCACTCTGTCGCCAGGCTGAGTGGCGTGGGATCTCAGATCAGTACCACT 84382
QY 63 CCATCTCTCGGGTCAATAATCTCTGCTCAGCCTCTCAAGTAGCTGGAGCTGTAG 122
Db 84383 CCGCCTCTCGGGTCAAGCAATCTCTGCTCAGCCTCGAGTCCGAGTAGCTGGAGCT 84437
QY 123 CTTCTAAGTAGCTGGGACACACCATGCCAGCTAAATTTTGTGTTTGTGTTTGT 182
Db 84438 -----ACAGGCACATGCCACCATACCCAGCTAAATTTTGTGTTTGTGTTTGT 84482
QY 183 AGAGCAGGTTTTCACCATGTTGGCCAGGCTGGTGCAGCTCCCTTTAGATCTCTCAT 242
Db 84483 AGAGCAGGTTTTCACCATGTTGGCCAGGCTGGTGCAGCTCCCTTTAGATCTCTCAT 84542
QY 243 CTGCTCTATCTCTTCCCTTCTTAATGCAGTATCCAGTTTCTTACTTATC--ACATTTA 300
Db 84543 CCGCCTTGCTTCCCAAGTCTGGGATTACAGGCTGAGCCAGGTCGCCGCTGGTC 84602
QY 301 TTATTTATCTTATTTATTTAGACAGAGCTTTGCTTTGTCGCCAAGGCTGGAGTACAGT 360
Db 84603 ATGTATTTCTCTTTTATTTTATAGAGGAGTCTGCTGCTGCTGCCAGGCTGGAGTACAGT 84662
QY 361 GGTGGAGTCTGGCTCAGTGCAGCTCCACCTGCTGGGTTTACAGCCATCTTCCGCCCTCA 420
Db 84663 GGGCGGATCTGAGTCTCAGTGCAGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 84722
QY 421 GCCTCCCCAGTAGCTGGGAGTAAAGCGCTGCCACAGCCCGCTAAATTTTGTAT 480
Db 84723 GCCTCCGAGTAGCTGGGAGTAAAGCGCTGCCACAGCCCGCTAAATTTTGTAT 84782
QY 481 TTTTATTAAGACGGGTTTC 501
Db 84783 TTTTATTAAGACGGGTTTC 84803

RESULT 8
AC024488/c
LOCUS
DEFINITION Homo sapiens chromosome 1 clone RP11-137A6 map 1, WORKING DRAFT
SEQUENCE, 20 unordered pieces.
ACCESSION AC024488
VERSION AC024488.3 GI:8954147
KEYWORDS HTG; HTGS_PHASE1; HTGS_DRAFT.
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini;
Hominoidea; Homo.
1 (bases 1 to 181655)
Birren,B., Linton,L., Nusbaum,C., Lander,E., Abraham,H., Allen,N.,
Anderson,S., Baldwin,J., Barna,N., Beda,F., Boquslavskiy,L.,
Boukhgalter,B., Brown,A., Burkett,G., Campopiano,A., Castle,A.,
Choepl,Y., Colangelo,M., Collins,S., Collymore,A., Cooke,P.,
Dearellano,K., Dewar,K., Diaz,J.S., Dodge,S., Domino,M., Doyle,M.,
Fenesstor,J., Ferreira,P., FitzHugh,W., Forrest,C., Gage,D.,
Galagan,J., Gardyna,S., Ginde,S., Goyette,M., Graham,L.,
Grand-Pierre,N., Grant,G., Hagos,B., Heaford,A., Horton,L.,

```

Howland,J.C., Iliev,I., Johnson,R., Jones,C., Kann,L., Karatas,A., Klein,J., Landers,T., Larcocque,K., Lehoczyk,J., Levine,R., Liu,C., Liu,G., Locke,K., Macdonald,P., Marquis,N., McCarthy,M., McEwan,P., McGurk,A., McKernan,K., McPheeters,R., Meldrum,J., Menes,L., Mihova,T., Miranda,C., Mienga,V., Morrow,J., Naylor,J., Norman,C.H., O'Connor,T., O'Donnell,P., O'Neill,D., Olivaz,T.M., Peterson,K., Pierre,N., Pisan,C., Pollara,V., Raymond,C., Riley,R., Rogov,P., Rothman,D., Roy,A., Santos,R., Schauer,S., Severy,P., Spencer,B., Stange-Thomann,N., Stojanovic,N., Subramanian,A., Talamas,J., Tesfaye,S., Theodore,J., Tirrell,A., Travers,M., Trigglio,J., Vassiliev,H., Viel,R., Vo,A., Wilson,B., Wu,X., Wyman,D., Ye,W.J., Young,G., Zainoun,J., Zimmer,A. and Zody,M.

Direct Submission
Submitted (28-FEB-2000) Whitehead Institute/MIT Center for Genome Research, 320 Charles Street, Cambridge, MA 02141, USA
On Jul 7, 2000 this sequence version replaced gi:7249343.
All repeats were identified using RepeatMasker:
Smit,A.F.A. & Green, P. (1996-1997)
http://ftp.genome.washington.edu/RM/RepeatMasker.html

----- Genome Center
Center: Whitehead Institute/ MIT Center for Genome Research
Center code: WIBR
Web site: http://www-seq.wi.mit.edu
Contact: sequence_submissions@genome.wi.mit.edu

----- Project Information
Center project name: L7151
Center clone name: 137_A_6

----- Summary Statistics
Sequencing vector: M13; M77815; 100% of reads
Chemistry: Dye-terminator Big Dye; 100% of reads
Assembly program: Phrap; version 0.960731
Consensus quality: 167600 bases at least Q40
Consensus quality: 175084 bases at least Q30
Consensus quality: 178093 bases at least Q20
Insert size: 182000; agarose-fp
Insert size: 179755; sum-of-contigs
Quality coverage: 4.1 in Q20 bases; agarose-fp
Quality coverage: 4.2 in Q20 bases; sum-of-contigs

TITLE JOURNAL COMMENT

* NOTE: This is a 'working draft' sequence. It currently
* consists of 20 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.

* 1 1308: contig of 1308 bp in length
* 1309: gap of 100 bp
* 1409: contig of 1385 bp in length
* 2793: gap of 100 bp
* 2794: contig of 1095 bp in length
* 3988: gap of 100 bp
* 4089: contig of 1236 bp in length
* 5325: gap of 100 bp
* 5425: contig of 1603 bp in length
* 7027: gap of 100 bp
* 7127: contig of 1355 bp in length
* 8482: gap of 100 bp
* 8582: contig of 3282 bp in length
* 8843: gap of 100 bp
* 11865: contig of 3868 bp in length
* 11965: gap of 100 bp
* 15832: contig of 4196 bp in length
* 15932: gap of 100 bp
* 20128: contig of 5962 bp in length
* 20228: gap of 100 bp
* 26190: contig of 7895 bp in length
* 26291: gap of 100 bp
* 34185: contig of 8696 bp in length
* 34285: gap of 100 bp
* 42982: contig of 7079 bp in length
* 43082: gap of 100 bp
* 50160: contig of 7079 bp in length

```
* 50161 50260: gap of 100 bp
* 50261 60943: contig of 10683 bp in length
* 60944 61043: gap of 100 bp
* 61044 70244: contig of 9201 bp in length
* 70245 70345: gap of 100 bp
* 70345 81838: contig of 11494 bp in length
* 81839 90735: gap of 100 bp
* 90735 90835: contig of 8797 bp in length
* 90836 105643: gap of 100 bp
* 105643 123454: contig of 14807 bp in length
* 123454 123554: gap of 100 bp
* 123554 181655: contig of 17711 bp in length
* 181655 43082: contig of 58102 bp in length.
FEATURES
  source
    1..181655
      /organism="Homo sapiens"
      /mol_type="genomic DNA"
      /db_xref="taxon:9606"
      /chromosome="1"
      /map="1"
      /clone.lib="RPC1-11 Human Male BAC"
      /clone="RP11-137A6"
    1..1308
      /note="assembly_fragment"
    1309..1408
      /estimated_length=100
    1409..2793
      /note="assembly_fragment"
    2794..2893
      /estimated_length=100
    2894..3988
      /note="assembly_fragment"
    3989..4088
      /estimated_length=100
    4089..5324
      /note="assembly_fragment"
    5325..5424
      /estimated_length=100
    5425..7027
      /note="assembly_fragment"
    7028..7127
      /estimated_length=100
    7128..8482
      /note="assembly_fragment"
    8483..8582
      /estimated_length=100
    8583..11864
      /note="assembly_fragment"
    11865..11964
      /estimated_length=100
    11965..15832
      /note="assembly_fragment"
    15833..15932
      /estimated_length=100
    15933..20128
      /note="assembly_fragment"
    20129..20228
      /estimated_length=100
    20229..26190
      /note="assembly_fragment"
    26191..26290
      /estimated_length=100
    26291..34185
      /note="assembly_fragment"
    34186..34285
      /estimated_length=100
    34286..42981
      /note="assembly_fragment"
    42982..43081
      /estimated_length=100
    43082..50160
      /note="assembly_fragment"
```

```
gap 50161..50260
      /estimated_length=100
misc_feature 50261..60943
      /note="assembly_fragment"
gap 60944..61043
      /estimated_length=100
misc_feature 61044..70244
      /note="assembly_fragment"
gap 70245..70344
      /estimated_length=100
misc_feature 70345..81838
      /note="assembly_fragment"
gap 81839..81938
      /estimated_length=100
misc_feature 81939..90735
      /note="assembly_fragment"
      clone_end:SP6
      vector_side:left
gap 90736..90835
      /estimated_length=100
misc_feature 90836..105642
      /note="assembly_fragment"
gap 105643..105742
      /estimated_length=100
misc_feature 105743..123453
      /note="assembly_fragment"
      clone_end:T7
      vector_side:right
gap 123454..123553
      /estimated_length=100

Query Match 47.9%; Score 239.8; DB 14; Length 181655;
Best Local Similarity 72.3%; Pred. No. 5.9e-57;
Matches 362; Conservative 0; Mismatches 117; Indels 22; Gaps 3;

QY 3 GGGGCTCACTCTGTGCGCCAGCGTGGAGTGGGTGGCGGTGATCTCAGATCACTACAACCT 62
Db 120547 GAGTCTCACTCTGTGCGCCAGCGTGGAGTGGGTGATCTCAGATCACTACAACCT 120488

QY 63 CCATCTCCTGGGTTCAAAATATCTCTTGCCCTCAGCCTCTCAGTAGCTGGGACTGTAG 122
Db 120487 CCGCCTCCTGGGTTCAAGCAATTTCTCTGCGCTCAGCCTCCGAGTAGCTGGGACT----- 120433

QY 123 CTCCTAAGTAGCTGGGCACACACACCATGCGCCAGCTAAATTTTGTGTGTTTTTTTGGT 182
Db 120432 -----ACAGGCACATGCCACCATACCAGCTAAATTTTGG-----TATTTTAGT 120388

QY 183 AGAGACAGGTTTTACCATGTTGGCCAGGCTGGGTGACCTCCCTTTTAGATTTCTCCTCAT 242
Db 120387 AGAGACAGGTTTTACCATGTTAGCCAGGATGTTCTGAATCTCTGACCTTGTAAATCCAC 120328

QY 243 CCTGCTCTATTCTTCCCTTTCTAATGCAAGTATCCAGTTTCTTACTTATC--ACATTTA 300
Db 120327 CCGCCTTGCGCTTCCCAAAGTGTCTGGGATTAACAGGCGTGAGCCCGGCTGGGTC 120268

QY 301 TTATTTATTTTATTATTATTGACAGAGTCTTGTGTCGCCAAGGCTGGAGTACAGT 360
Db 120267 ATGATATTCTTTTATTTTATAGCGAGTCTTCGCTCTGTGCCAGGCTGGAGTGCAGT 120208

QY 361 GGTGGGATCTCGGCTCACTGCAAGCTCCAACCTGTGGGTTACGGCAATTTCTCCGCTCA 420
Db 120207 GGGCGGATCTCAGCTCACTGCAAGCTCTGCGCTCTCTGGGTTACGGCAATTTCTCTGCTCA 120148

QY 421 GCCTCCCCAGTAGCTGGGACTAAAGCGCGCTCGCCACCGCCCCCGCTAATTTTTTGTAT 480
Db 120147 GCCTCCCGAGTAGCTGGGACTACAGGTGCTCGCCACCGCCAGCTAATTTTTTGTAT 120088

QY 481 TTTTAAATAAGCGGGTTTC 501
Db 120087 TTTTAGTAGAGCGGGTTTC 120067

RESULT 9
```


AC116046

LOCUS AC116046 193352 bp DNA linear HTG 04-FEB-2004
 DEFINITION Papio hamadryas clone RP41-343P17, *** SEQUENCING IN PROGRESS ***
 4 ordered pieces.

ACCESSION

AC116046

VERSION

AC116046.2 GI:42270640

KEYWORDS

HTG; HTGS_PHASE2; HTGS_FULTOP; HTGS_ACTIVAFIN.

SOURCE

Papio hamadryas (hamadryas baboon)

ORGANISM

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini;

Cercopithecoidea; Cercopithecinae; Papio.

1 (bases 1 to 193352)

Birren,B., Nusbaum,C. and Lander,E.

Papio hamadryas, clone RP41-343P17

REFERENCE

AUTHORS

TITLE

JOURNAL

REFERENCE

AUTHORS

TITLE

JOURNAL

REFERENCE

AUTHORS

TITLE

JOURNAL

REFERENCE

AUTHORS

TITLE

JOURNAL

REFERENCE

AUTHORS

TITLE

JOURNAL

REFERENCE

AUTHORS

TITLE

JOURNAL

REFERENCE

AUTHORS

TITLE

JOURNAL

REFERENCE

AUTHORS

TITLE

JOURNAL

REFERENCE

AUTHORS

TITLE

JOURNAL

REFERENCE

AUTHORS

TITLE

JOURNAL

REFERENCE

AUTHORS

TITLE

JOURNAL

REFERENCE

AUTHORS

TITLE

JOURNAL

REFERENCE

AUTHORS

TITLE

JOURNAL

REFERENCE

AUTHORS

TITLE

JOURNAL

REFERENCE

AUTHORS

TITLE

JOURNAL

REFERENCE

AUTHORS

TITLE

Web site: <http://www-seq.wi.mit.edu>

Contact: sequence_submissions@genome.wi.mit.edu

----- Project Information

Center project name: L12531

Center clone name: 343_P_17

* NOTE: This is a 'working draft' sequence. It currently

* consists of 4 contigs. Gaps between the contigs

* are represented as runs of N. The order of the pieces

* is believed to be correct as given, however the sizes

* of the gaps between them are based on estimates that have

* provided by the submitter.

* This sequence will be replaced

* by the finished sequence as soon as it is available and

* the accession number will be preserved.

* 1 41116: contig of 41116 bp in length

* 41117 41216: gap of 100 bp

* 41217 127557: contig of 86341 bp in length

* 127558 127657: gap of 100 bp

* 127658 157700: contig of 30043 bp in length

* 157701 157800: gap of 100 bp

* 157801 193352: contig of 35552 bp in length.

* Location/Qualifiers

1.193352

/organism="Papio hamadryas"

/mol_type="genomic DNA"

/db_xref="taxon:9557"

/clone="RP41-343P17"

/Clone_lib="RPC1-41 Male Baboon BAC"

41117..41216

/estimated_length=100

127558..127657

/estimated_length=100

157701..157800

/estimated_length=100

ORIGIN

Query Match 47.6% Score 238.6; DB 14; Length 193352;

Best Local Similarity 70.7% Pred No. 1.3e-56;

Matches 353; Conservative 0; Mismatches 129; Indels 17; Gaps 2;

Qy 3 GGGGCTCAGTCTGTGCGCCAGGCTGGAGTGGGCTGATCTCAGATCACTACACCT 62

Db 114678 GAGTCTCGCTCTGTAGCCAGGCTGGAGTGGGCTGATCTCAGTCACTACACCT 114737

Qy 63 CCATCTCTGGTTCATAAATTAATCTCTGCTGCTCAGCTCTCAAGTAGCTGGAGCTTAG 122

Db 114738 CCGCTCTCCGGGTTACGCCATTTCTCGGCTCAGCTCCGAGTAGCTGGGACT----- 114792

Qy 123 CTCTCAAGTAGCTGGGACACACACCATGCCCCAGCTAAATTTTGTGTGTTTTTGGT 182

Db 114793 -----ACAGGGCGCGCCACCTCGCGCTGATTTTCTTTTCTTAGT 114842

Qy 183 AGAGACAGGTTTTTACCATGTTGGCCAGGCTGGGCTGACCTCTTTTGTAGATTTCTCTCAT 242

Db 114843 AAAGACGGGTTTCCCGTGTAGCCAGGATGG--TCTCGATCTCTGACCTCGTGATCC 114900

Qy 243 CTTGCTCTATTCTTCCCTTTCTAATGACGATTCAGTTTCTTCTTATACATTTATT 302

Db 114901 GCCCATCTCGGCTCCCAAAGTCTGGGATACAGGCTTGAGCCACCGCGCGCTGA 114960

Qy 303 ATTATTCTTATTATTATTTAGACAGAGTCTTGTGTTGCGCCAGGCTGGAGTACAGTGG 362

Db 114961 CTTTTTTTTTTTTTTTTTGTAGACGGAGTCTCGCGCTGTGCGCCAGGCTGGAGTACAGTGG 115020

Qy 363 TGGGATCTCGGCTCACTCAAGCTCCACCTGTGGGTTTACGCGCATTTCTCCGCTCAGC 422

Db 115021 CGGATCTTGGCTCACTGCAAGCTCCGCTCCCGGGTTTACGCGCATTTCTCTGCTCAGC 115080

Qy 423 CTCCGAGTAGCTGGGACTAAAGCGCTGCCACACCGCCCGCTTAATTTTTTGTATTT 482

Db 115081 CTCTGAGTAGCTGGTACTACAGCGCCCGCCACCGCGCGCTAATTTTTTGTATTT 115140

MAPPING INFORMATION:
The source of this clone was established as part of a mapping an-

```

region
4752..5441
/db_xref="dbSNP:953770"
/allele="C"
/allele="T"

```

repeat region

```

repeat_region /rpt_family="ERV1"
5526..5656
repeat_region /rpt_family="Alu"
5769..6067
repeat_region /rpt_family="Alu"
6154..6190
repeat_region /rpt_family="Alu"
6220..6500
repeat_region /rpt_family="Alu"
7547..7687
repeat_region /rpt_family="MIR"
7779..7961
repeat_region /rpt_family="Alu"
7965..8266
repeat_region /rpt_family="Alu"
9058..10603
misc_feature /note="CpG island (%GC=66.6, o/e=0.99, #CpGs=168)"
9535..9607
repeat_region /rpt_family="GC-rich"
9916
variation /allele="C"
/allele="G"
/allele="T"
variation /db_xref="dbSNP:2301943"
/allele="C"
/allele="G"
variation /db_xref="dbSNP:2301942"
10600
/allele="C"
/allele="A"
/allele="G"
repeat_region /db_xref="dbSNP:2398661"
10697..10790
repeat_region /rpt_family="Alu"
10821..11015
repeat_region /rpt_family="L1"
11016..11148
repeat_region /rpt_family="Alu"
11149..11182
repeat_region /rpt_family="L1"
11189..11375
repeat_region /rpt_family="Alu"
11380..11412
repeat_region /rpt_family="L1"
11414..11581
repeat_region /rpt_family="Alu"
11672..11744
repeat_region /rpt_family="MIR"
11745..12044
repeat_region /rpt_family="Alu"
12048..12282
repeat_region /rpt_family="MER2_type"
12283..12383
repeat_region /rpt_family="waLR"
12386..12691
repeat_region /rpt_family="Alu"
12720..13046
repeat_region /rpt_family="Alu"
13049..13321
repeat_region /rpt_family="Alu"
13461..13655
repeat_region /rpt_family="Alu"

Query Match 47.3%; Score 237; DB 8; Length 123253;
Best Local Similarity 72.1%; Pred. No. 3.6e-56;
Matches 360; Conservative 0; Mismatches 115; Indels 24; Gaps 3;

Qy 3 GGGGCTCACTGTGGCCAGGCTGGAGTGGCGTGGCTGATCTCAGATCACTACACCT 62
Db 28501 GAGTTTCACTGTCAACCCAGGCTGGAGTGGCGTGGCTGATCTGATCTCACTCACT 28560

Qy 63 CCATCTCTGGGTTCAATATATCTTCTGCTCAGCCCTCTCAAGTAGCTGGACTTAG 122
Db 28561 CTGGCTCCAGGGTTCAACAAATCTCTGTCTCAGCCTCTCAAGTATCTGGGA----- 28613

```

```

Qy 123 CTCTCAAGTAGCTGGCACACACACCATGCCAGCTAATTTTGTGTGTTTTTTTGTGT 182
Db 28614 -----TGACAGAGGCTGCACACCATGCCAGCTAATTTTAG-----CAATTTTAGT 28660
Qy 183 AGAGACAGGTTTTCACCATGTTGGCAGGCTGGGTGACCTCCCTTTTATAGATCTCTCAT 242
Db 28661 AGAGATGGGTTTTCACCATGTTGGCAGGCTGGCATTAATCTGATTTCTTTCCACCAAA 28720
Qy 243 CCTGCTCTATTTCTTCCCGCTTCTAATGCAGTATCCAGTTTCTTACTTATCACAATTTAT 302
Db 28721 TTAACATTCATTT-----TCTGAATTCATCAAGTTTTCCTACTGTAGAATTTTTTT 28776
Qy 303 ATTATTTCTTATTTATTTGACAGAGTCTTGTCTTGTTCGCAAGGCTGGAGTACAGTGG 362
Db 28777 TTTTCTTTTCTTTTGTGACAGAGTCTGCTCTGCTGCGCCAGGCTGGAGTGCAGTGC 28836
Qy 363 TGGCATCTGGCTCACTGCAAGCTCCACCTGCTGGTTCACGCCATTCCTCGGCTCAGC 422
Db 28837 CGCGTCTCGGCTCACCGCAAGCTCCGCTCCAGGTTTCAGGCCATTCGCTGCTCAGC 28896
Qy 423 CTCCCCAGTAGCTGGGACTAAGGCGCTGCACACGCGCGCTAATTTTCTTATTTTCTATT 482
Db 28897 CTCCCAAGTGGCTGGGACTACAGGCGCTGCACACGCGCGCTAATTTTCTTATTTTCTATT 28956
Qy 483 TTAATAAAGACGGGGTTTC 501
Db 28957 TTAGTAGAGAGGGGTTTC 28975

```

RESULT 11 AC005666/c

LOCUS AC005666 158905 bp DNA linear PRI 22-DEC-1998
DEFINITION Homo sapiens chromosome 17, clone hRPK.112_H_10, complete sequence.
ACCESSION AC005666
VERSION AC005666.1 GI:4049331
KEYWORDS HTG.
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini;
Hominidae; Homo.
REFERENCE 1 (bases 1 to 158905)
AUTHORS Birren,B., Linton,L., Nusbaum,C. and Lander,E.
TITLE Homo sapiens chromosome 17, clone hRPK.112_H_10
JOURNAL Unpublished
REFERENCE 2 (bases 1 to 158905)
AUTHORS Birren,B., Fasman,K., Linton,L., Nusbaum,C., Lander,E., Allen,N.,
Anderson,M., Baker,J., Baldwin,J., Barna,N., Beckerly,R., Benn,J.,
Boutwell,C., Brown,A., Castle,A., Cerny,J., Colangelo,M.,
Collins,S., Collymore,A., Cooke,P., Corliss,D., Depayre,E.,
Devon,K., Dewar,K., Doneelan,L., Ferreira,P., Fitzhugh,W.,
Forrest,C., Funke,R., Gage,D., Gardyna,S., Geraghty,K., Grant,G.,
Hagos,B., Heaford,A., Herena,L., Horton,L., Howland,J.C.,
Jacotot,L., Jones,C., Kann,L., Karatas,A., Lehocsky,J.,
Macdonald,P., Marquis,N., McEwan,P., McGurk,A., McKernan,K.,
Meldrim,J., Molla,M., Morris,W., Morrow,J., Mychaleckyj,J.,
Nahf,R., Naylor,J., Niloff,M., O'Connor,I., O'Donnell,P.,
Pavlin,B., Peterson,K., Riley,R., Roberts,D., Roy,A., Severy,P.,
Stange-Thomann,N., Stilwell,J., Stojanovic,N., Stone,C.,
Subramanian,A., Tesfaye,S., Tichovolsky,N., Torruella-Miller,I.,
Vassiliev,H., Vo,A., Wagner,A., Wheeler,J., Wu,Y., Wyman,D.,
Ye,W.J., Zhao,J. and Zody,M.

Direct Submission
TITLE
JOURNAL
Submitted (10-SEP-1998) Whitehead Institute/MIT Center for Genome
Research, 320 Charles Street, Cambridge, MA 02141, USA
REFERENCE 3 (bases 1 to 158905)
AUTHORS Birren,B., Linton,L., Nusbaum,C., Lander,E., Allen,N., Anderson,M.,
Baker,J., Baldwin,J., Barna,N., Beckerly,R., Benn,J., Brown,A.,
Castle,A., Cerny,J., Colangelo,M., Collins,S., Collymore,A.,
Cooke,P., Dearellano,K., Depayre,E., Devon,K., Dewar,K.,
Doneelan,L., Ferreira,P., Fitzhugh,W., Forrest,C., Funke,R.,
Gage,D., Galagan,J., Gardyna,S., Grant,G., Hagos,B., Heaford,A.,

Herena, L., Horton, L., Howland, J.C., Jones, C., Kann, L., Karatas, A.,
 Lehoczy, J., Macdonald, P., Marquis, N., McEwan, P., McGurk, A.,
 McKernan, K., Meltrin, J., Molla, M., Morris, W., Morrow, J.,
 Mychaleckyj, J., Naylor, J., Niloff, M., O'Connor, T., O'Donnell, P.,
 Pavlin, B., Peterson, K., Pollara, V., Riley, R., Roberts, D., Roy, A.,
 Severy, P., Stange-Thomann, N., Stojanovic, N., Stone, C.,
 Subramanian, A., Testaye, S., Torruella-Miller, I., Vassiliev, H.,
 Vo, A., Wagner, A., Wheeler, J., Wu, X., Wyman, D., Ye, W.-J. and Zody, M.
 Direct Submission
 Submitted (22-DEC-1998) Whitehead Institute/MIT Center for Genome
 Research, 320 Charles Street, Cambridge, MA 02141, USA
 On Dec 22, 1998 this sequence version replaced gi:4028945.
 All repeats were identified using RepeatMasker: Smit, A.F.A. &
 Green, P. (1996-1997)
<http://ftp.genome.washington.edu/RM/RepeatMasker.html>.

FEATURES

Location/Qualifiers

source

1..158905
 /organism="Homo sapiens"
 /mol_type="genomic DNA"
 /db_xref="taxon:9606"
 /chromosome="17"
 /map="17"
 /clone="hRPK.112.H.10"
 /clone_lib="RPGI-II human BAC library"

repeat_region
 complement(835..1135)
 /rpt_family="AluSc"
 repeat_region
 complement(1624..1918)
 /rpt_family="AluSg"
 repeat_region
 2302..2467
 /rpt_family="L1MB3"
 repeat_region
 2471..2778
 /rpt_family="AluY"
 repeat_region
 2779..2925
 /rpt_family="AluJo/FRAM"
 repeat_region
 2926..2948
 /rpt_family="CAAA)n"
 repeat_region
 2955..3255
 /rpt_family="L1MB3"
 repeat_region
 3256..3560
 /rpt_family="AluSc"
 repeat_region
 3570..3729
 /rpt_family="FAM"
 repeat_region
 3730..3787
 /rpt_family="L1MB3"
 repeat_region
 complement(3964..4114)
 /rpt_family="MERSB"
 repeat_region
 5111..5426
 /rpt_family="AluJb"
 repeat_region
 5480..5515
 /rpt_family="AT_rich"
 repeat_region
 5752..5784
 /rpt_family="AT_rich"
 repeat_region
 complement(5999..6328)
 /rpt_family="MERSB"
 repeat_region
 6435..6518
 /rpt_family="L1MB5"
 repeat_region
 6519..6815
 /rpt_family="AluY"
 repeat_region
 6816..7533
 /rpt_family="L1MB5"
 repeat_region
 7535..7694
 /rpt_family="AluJb"
 repeat_region
 7705..8001
 /rpt_family="AluSg"
 repeat_region
 8124..8183
 /rpt_family="MLT2D"
 repeat_region
 complement(8184..8205)
 /rpt_family="AluSc"
 repeat_region
 complement(8206..8229)
 /rpt_family="CAAA)n"
 repeat_region
 complement(8230..8515)
 /rpt_family="AluSg"
 repeat_region
 8516..8942

repeat_region
 /rpt_family="MLT2D"
 complement(8957..9079)
 /rpt_family="MER7A"
 9080..9419
 /rpt_family="THE1C"
 complement(9420..9626)
 /rpt_family="MER7A"
 9702..9748
 /rpt_family="MLT2D"
 9863..9940
 /rpt_family="L2"
 10022..10238
 /rpt_family="L1MB3"
 complement(10325..11142)
 /rpt_family="L2"
 complement(11143..11173)
 /rpt_family="TGAA)n"
 complement(11175..11440)
 /rpt_family="AluJb"
 complement(11441..12055)
 /rpt_family="L2"
 complement(12056..12364)
 /rpt_family="AluSg"
 complement(12365..12447)
 /rpt_family="L2"
 12448..12722
 /rpt_family="AluSc"
 12726..13024
 /rpt_family="AluJb"
 complement(13025..13417)
 /rpt_family="L2"
 13419..13698
 /rpt_family="AluSc"
 13699..13726
 /rpt_family="TAAA)n"
 13783..14076
 /rpt_family="AluY"
 complement(14083..14322)
 /rpt_family="MIR"
 14484..14510
 /rpt_family="AT_rich"
 complement(15900..16212)
 /rpt_family="L1MA4A"
 16280..16596
 /rpt_family="MER7A"
 16986..17306
 /rpt_family="AluJo"
 17422..17468
 /rpt_family="CAAAA)n"
 17550..17642
 /rpt_family="HAL1"
 complement(17643..17937)
 /rpt_family="AluSp"
 17938..18133
 /rpt_family="HAL1"
 18764..19102
 /rpt_family="L1ME"
 19121..19742
 /rpt_family="L1M4"
 20318..20493
 /rpt_family="MLT13"
 complement(20498..20831)
 /rpt_family="L1M4"
 complement(20888..21059)
 /rpt_family="AluSg"
 complement(21060..21328)
 /rpt_family="AluY"
 complement(21329..21460)
 /rpt_family="AluSg"
 22123..23454
 /rpt_family="L1MB2"
 23455..23586
 /rpt_family="FLAM_C"

Quality coverage: 7.5 in Q20 bas.
* NOTE: This is a 'working draft' sequence. It currently
* consists of 14 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.

* 1 6481: contig of 6481 bp in length
* 6482 6581: gap of 100 bp
* 6582 6995: contig of 414 bp in length
* 6996 7095: gap of 100 bp
* 7096 47880: contig of 40785 bp in length
* 47881 49647: contig of 100 bp
* 49648 49747: contig of 1667 bp in length
* 49748 54536: contig of 4789 bp in length
* 54537 54636: gap of 100 bp
* 54637 58724: contig of 4088 bp in length
* 58725 63405: gap of 100 bp
* 63406 63505: contig of 4581 bp in length
* 63506 70593: gap of 100 bp
* 70594 83758: contig of 13065 bp in length
* 83759 94549: contig of 10691 bp in length
* 94550 94649: gap of 100 bp
* 94650 109171: contig of 14522 bp in length
* 109172 109271: gap of 100 bp
* 109272 126036: contig of 16765 bp in length
* 126037 126136: gap of 100 bp
* 126137 170138: contig of 44002 bp in length
* 170139 170238: gap of 100 bp
* 170239 170892: contig of 654 bp in length.
* 170239 170892: contig of 654 bp in length.

FEATURES

source
1. .170892
/organism="Homo sapiens"
/mol_type="genomic DNA"
/db_xref="taxon:9606"
/chromosome="17"
/map="17"
/clone="RP11-110F1"
/clone_lib="RPC1-11 Human Male BAC"
1. .6481
/note="assembly_fragment
clone_end:SF6
vector_side:left"
6482. .6581
/estimated_length=100
6582. .6995
/note="assembly_fragment"
6996. .7095
/estimated_length=100
7096. .47880
/note="assembly_fragment"
47881. .47980
/estimated_length=100
47981. .49647
/note="assembly_fragment"
49648. .49747
/estimated_length=100
49748. .54536
/note="assembly_fragment"
54537. .54636
/estimated_length=100
54637. .58724
/note="assembly_fragment"
58725. .58824
/estimated_length=100
58825. .63405
/note="assembly_fragment"

gap 63406. .63505
/estimated_length=100
misc_feature 63506. .70593
/note="assembly_fragment"
70594. .70693
/estimated_length=100
misc_feature 70694. .83758
/note="assembly_fragment"
83759. .83858
/estimated_length=100
misc_feature 83859. .94549
/note="assembly_fragment"
94550. .94649
/estimated_length=100
misc_feature 94650. .109171
/note="assembly_fragment"
109172. .109271
/estimated_length=100
misc_feature 109272. .126036
/note="assembly_fragment"
126037. .126136
/estimated_length=100
misc_feature 126137. .170138
/note="assembly_fragment"
170139. .170238
/estimated_length=100
misc_feature 170239. .170892
/note="assembly_fragment
clone_end:T7
vector_side:right"
ORIGIN
Query Match 47.3%; Score 236.8; DB 14; Length 170892;
Best Local Similarity 73.0%; Pred. No. 4.2e-56;
Matches 371; Conservative 0; Mismatches 107; Indels 30; Gaps 4;
QY 5 GGCTCACTCTCTGCCCCAGGCTGGAGTGGCGTGGCGTATCTCAGATCACTACAACTCC 64
Db 120205 GTCTCACTCTCTCACCAGGCTGGAGTGCATGGCGGATCTTGGCTCACTGCAAACTCT 120146
QY 65 ATCTCTCGGTTCAAATAATTCTCTTGGCTCAGCCTCTCAAGTAGCTGGGACTTGTAGCT 124
Db 120145 GCCTCTCGGTTCAAATGATTTCTCGTCTCAGCCTCCCAAGTAGCTGGGA----- 120095
QY 125 CTCAGTAGCTGGGCACACACCACCATGCCAGCTAAATTTTGTGTTTTTTTGGTAG 184
Db 120094 -----TTACAGGTACAGCCACCAAGCCAGCTAAATTTTTT---TGTATTTTTTGGTAG 120044
QY 185 AGACAGGTTTTTCAACCATGTTGGCCAGGCTGGGTGACCTCCC-----TTTTAGAT 233
Db 120043 AAACAGAAATTTTCAACCATGTTGGCCAGGCTGCTCAAACTCTGACCTCAAGTGATCTGC 119984
QY 234 TCTCTCATCTCTGCTCTATTCTTCCCTTTCTTAATGCGAGTATCCAGTTTCTTACTTATC 293
Db 119983 CTGCTCTGGCTCTCCACAGTGTCTAGGATTTACAGGTGTGAGCTACCGTGTCCAGCTATGC 119924
QY 294 ACATTTATTTATTTCTTTATTTATTTATTTATTTATTTATTTATTTATTTATTTATTT 353
Db 119923 ATATCTTGAGACGGAGTCTCGCTCTGTACCCAGGCTGGA 119864
QY 354 GTACAGTGGTGGGATCTGGCTCACTGCAAGCTCCACCTGCTGGGTTTTCAGGCAATTTCTCC 413
Db 119863 GTGCGAGTGGGGGATCTCGGCTCACTGCAAGCTCTCGTCTCCCGGGTTCAAGCAATTTCTCC 119804
QY 414 CGCCTCAGCCTCCCGAGTAGCTGGGACTAAAGCGCCTGCCACCGCCCGCGCTAAATTTT 473
Db 119803 TGGCTCAGCCTCCCAAGTAGCTGGGACTACAGCGCGCCGCCACTACCGCCGGCTAA-TTT 119745
QY 474 TTTGTTATTTTAAATAAGACGGGGTTTC 501
Db 119744 TTTGTTATTTTATAGTAGAGACGGGGTTTC 119717

RESULT 13

AC099811/c

LOCUS

DEFINITION

AC099811

AC099811

VERSION

KEYWORDS

SOURCE

ORGANISM

AC099811 174902 bp DNA linear PRI 07-FEB-2003
Homo sapiens chromosome 17, clone RP11-358B23, complete sequence.

AC099811

HTG. AC099811.7 GI:28269667

Homo sapiens (human)

Homo sapiens

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini;

Hominidae; Homo

1 (bases 1 to 174902)

Birren,B., Nusbaum,C. and Lander,E.

Homo sapiens chromosome 17, clone RP11-358B23

Unpublished

2 (bases 1 to 174902)

Birren,B., Linton,L., Nusbaum,C., Lander,E., Ali,A., Allen,N.,

Anderson,S., Barna,N., Bastien,V., Boguslavskiy,L., Boukhgalter,B.,

Brown,A., Camarata,J., Campopiano,A., Chang,J., Chazaro,B.,

Choepel,Y., Colangelo,M., Collins,S., Collamore,A., Cook,A.,

Cooke,P., DeArrellano,K., Dewar,K., Diaz,J.S., Dodge,S., Faro,S.,

Ferreira,P., FitzGerald,M., Gage,D., Galagan,J., Gardyna,S.,

Ginde,S., Gord,S., Goyette,M., Graham,L., Grand-Pierre,N.,

Hagos,B., Hearford,A., Horton,L., Hulme,W., Iliev,I., Johnson,R.,

Jones,C., Kamat,A., Karatas,A., Kells,C., LaRocque,K.,

Lamazares,R., Landers,T., Lehoczy,J., Levine,R., Liu,G.,

MacLean,C., Macdonald,P., Major,J., Marquis,N., Matthews,C.,

McCarthy,M., McEwan,P., McKernan,K., McPheeters,R., Meldrim,J.,

Meneus,L., Mihova,T., Mienga,V., Murphy,T., Naylor,J., Nguyen,C.,

Norbu,C., Norman,C.H., O'Connor,T., O'Donnell,P., O'Neill,D.,

Oliver,J., Peterson,K., Phunkhang,P., Pierre,N., Pollara,V.,

Ramond,C., Retta,R., Rieback,M., Riley,R., Rise,C., Rogov,P.,

Roman,J., Rosetti,M., Roy,A., Santos,R., Schauer,S., Schuback,R.,

Seaman,S., Severy,P., Spencer,B., Stange-Thomann,N., Stojanovic,N.,

Straus,N., Subramanian,A., Talamas,J., Tesfaye,S., Theodore,J.,

Topham,K., Travers,M., Travis,N., Trigilio,J., Vassiliev,H.,

Viel,R., Vo,A., Wilson,B., Wu,X., Wyman,D., Ye,W.J., Young,G.,

Zainoun,J., Zembek,L., Zimmer,A. and Zody,M.

Direct Submission

Submitted (22-NOV-2001) Whitehead Institute/MIT Center for Genome

Research, 320 Charles Street, Cambridge, MA 02141, USA

3 (bases 1 to 174902)

Birren,B., Nusbaum,C., Lander,E., Ali,A., Allen,N., Anderson,S.,

Barna,N., Bastien,V., Bloom,T., Boguslavskiy,L., Boukhgalter,B.,

Camarata,J., Chang,J., Chazaro,B., Choepel,Y., Collamore,A.,

Cook,A., Cooke,P., DeArrellano,K., Dewar,K., Diaz,J.S., Dodge,S.,

Faro,S., Ferreira,P., FitzGerald,M., Gage,D., Galagan,J.,

Gardyna,S., Gord,S., Graham,L., Grand-Pierre,N., Hafez,N.,

Kamat,A., Karatas,A., Kells,C., Landers,T., Levine,R.,

Lindblad-Toh,K., Liu,G., MacLean,C., Macdonald,P., Major,J.,

Mathews,C., McCarthy,M., Meldrim,J., Meneus,L., Mihova,T.,

Mlenga,V., Murphy,T., Naylor,J., Nguyen,C., Nicol,R., Norbu,C.,

Norman,C.H., O'Connor,T., O'Donnell,P., O'Neill,D., Oliver,J.,

Peterson,K., Phunkhang,P., Pierre,N., Raymond,C., Retta,R.,

Rise,C., Rogov,P., Roman,J., Roy,A., Schauer,S., Schuback,R.,

Seaman,S., Severy,P., Smith,C., Spencer,B., Stange-Thomann,N.,

Stojanovic,N., Talamas,J., Tesfaye,S., Theodore,J., Topham,K.,

Travers,M., Vassiliev,H., Viel,R., Vo,A., Wilson,B., Wu,X.,

Wyman,D., Young,G., Zainoun,J., Zembek,L., Zimmer,A. and Zody,M.

Direct Submission

Submitted (14-JAN-2003) Whitehead Institute/MIT Center for Genome

Research, 320 Charles Street, Cambridge, MA 02141, USA

4 (bases 1 to 174902)

Birren,B., Nusbaum,C., Lander,E., Abouelleil,A., Allen,N.,

Anderson,S., Arachchi,H.M., Barna,N., Bastien,V., Bloom,T.,

Boguslavskiy,L., Boukhgalter,B., Camarata,J., Chang,J., Choepel,Y.,

Collamore,A., Cooke,A., Cooke,P., Corum,B., DeArrellano,K.,

Diaz,J.S., Dodge,S., Dooley,K., Dorris,L., Erickson,J., Faro,S.,

Ferreira,P., FitzGerald,M., Gage,D., Galagan,J., Gardyna,S.,

Graham,L., Grand-Pierre,N., Hafez,N., Hagopian,D., Hagos,B.,

Hall,J., Horton,L., Hulme,W., Iliev,I., Johnson,R., Jones,C.,

Kamat,A., Karatas,A., Kells,C., Landers,T., Levine,R.,
Lindblad-Toh,K., Liu,G., Lui,A., Mabbitt,R., MacLean,C.,
Macdonald,P., Major,J., Manning,J., Matthews,C., McCarthy,M.,
Meldrim,J., Meneus,L., Mihova,T., Mienga,V., Murphy,T., Naylor,J.,
Nguyen,C., Nicol,R., Norbu,C., O'Connor,T., O'Donnell,P.,
O'Neill,D., Oliver,J., Peterson,K., Phunkhang,P., Pierre,N.,
Rachupka,A., Ramasamy,U., Raymond,C., Retta,R., Rise,C., Rogov,P.,
Roman,J., Schauer,S., Schuback,R., Seaman,S., Severy,P., Smith,C.,
Spencer,B., Stange-Thomann,N., Stojanovic,N., Stubbs,M.,
Talamas,J., Tesfaye,S., Theodore,J., Topham,K., Travers,M.,
Vassiliev,H., Venkataraman,V.S., Viel,R., Vo,A., Wilson,B., Wu,X.,
Wyman,D., Young,G., Zainoun,J., Zembek,L., Zimmer,A. and Zody,M.

TITLE

JOURNAL

COMMENT

COMMENT

COMMENT

COMMENT

COMMENT

COMMENT

COMMENT

COMMENT

COMMENT

COMMENT

COMMENT

COMMENT

COMMENT

COMMENT

COMMENT

COMMENT

COMMENT

COMMENT

COMMENT

COMMENT

COMMENT

COMMENT

COMMENT

COMMENT

COMMENT

COMMENT

COMMENT

COMMENT

COMMENT

COMMENT

COMMENT

COMMENT

COMMENT

COMMENT

COMMENT

COMMENT

COMMENT

COMMENT

COMMENT

COMMENT

COMMENT

COMMENT

COMMENT

COMMENT

COMMENT

COMMENT

COMMENT

COMMENT

COMMENT

COMMENT

COMMENT

COMMENT

COMMENT

COMMENT

COMMENT

COMMENT

COMMENT

COMMENT

COMMENT

COMMENT

COMMENT

COMMENT

COMMENT

COMMENT

COMMENT

COMMENT

COMMENT

repeat_region	8766..9069	/rpt_family="AluSx"	183	AGAGACAGGTTTTCACCATGTTGGCCAGGCTGGGTGACCTCCCTCTTTAGATTCTCTCAT 242
repeat_region	9081..9200	/rpt_family="L1MB3"	30409	GGAGACAGGTTTTCACCATGTTGGCCAGGAT-GGTCTCAATCTCTGACCTCTGATCGG 30351
repeat_region	9348..9652	/rpt_family="AluSx"	243	CCTGCTCTATTCTTCCCTTTCTAAATGTCAGTATCCAGTTTCTTTACTTATCACAATTATT 302
repeat_region	complement(11210..11443)	/rpt_family="AluSx"	30350	CCTGCTCTCTCTCCCAAGTGTGGGATTTACAAGTGTGACGCCACCGCACCCAGCCTATA 30291
repeat_region	12226..12257	/rpt_family="AluY"	303	ATTAT-TCATTATTATTAGACAGAGTCTTGTGTCGCAAGGCTGGAGTACAGTG 361
repeat_region	complement(12258..12566)	/rpt_family="AluSx"	30290	AAAATATTTTTTTTTTTTGGAGCGGAGTCTCACTCTGTGCGCCAGGCTGGAGTGCAGTG 30231
repeat_region	complement(12567..12866)	/rpt_family="AluJb"	362	GTGCGATCTCGGCTCACTGCAAGCTCCACCTGTGGGTTTCAACGCAATTTCTCCGCCCTCAG 421
repeat_region	13359..13584	/rpt_family="MIR"	30230	GTGCAATCTCGGCTCACTGCAAACTCGGCTTCCGGGTTTCAAGCAATTTCTCTGCTCAG 30171
repeat_region	complement(13585..14075)	/rpt_family="AluJb"	422	CCTCCCGAGTGTGGGACTTAAAGGCGCTGCCACCAAGGCTGGGCTTAATTTTGTGATT 481
repeat_region	15039..15095	/rpt_family="L1MC3"	30170	CCTCCCGAGTGTGGGACTTAAAGGCGCTGCCACCAAGGCTGGGCTTAATTTTGTGATT 30111
repeat_region	15096..15381	/rpt_family="AluSx"	482	TTTAATAAGACGGGTTTC 501
repeat_region	15382..15524	/rpt_family="L1MC3"	30110	TTTAGTAGACGGGTTTC 30091
repeat_region	complement(15525..15826)	/rpt_family="AluSx"	RESULT 14	
repeat_region	15827..15915	/rpt_family="L1MC3"	AL357936	107529 bp DNA linear PRI 18-MAY-2005
repeat_region	15916..16226	/rpt_family="AluSx"	LOCUS	Human DNA sequence from clone RP11-524G24 on chromosome 9. Contains
repeat_region	complement(16227..16271)	/rpt_family="AluSx"	DEFINITION	the 5' end of two variants of the DAB21P gene for DAB2 interacting
repeat_region	16272..16386	/rpt_family="AluY"	ACCESSION	protein (AF9034, D1P1/2, KIAA1743), a novel gene and two Cpg
repeat_region	16392..16721	/rpt_family="AluY"	VERSION	islands, complete sequence.
repeat_region	16722..16939	/rpt_family="AluY"	KEYWORDS	AL357936.25 GI:18252285
repeat_region	16943..17252	/rpt_family="L1MC3"	SOURCE	HTG; AF9034; Cpg island; DAB21P; D1P1/2; KIAA1743.
repeat_region	17253..17507	/rpt_family="AluSx"	ORGANISM	Homo sapiens (human)
repeat_region	complement(17536..17847)	/rpt_family="AluSx"	REFERENCE	1 (bases 1 to 107529)
repeat_region	17848..18159	/rpt_family="L1MC3"	AUTHORS	Kimberley, A.
repeat_region	18161..18209	/rpt_family="AluJo"	TITLE	Direct Submission
repeat_region	18212..18506	/rpt_family="AluSx"	JOURNAL	Submitted (13-MAY-2005) Wellcome Trust Sanger Institute, Hinxton,
repeat_region	18507..18738	/rpt_family="AluSx"	COMMENT	Cambridgeshire, CB10 1SA, UK. E-mail enquiries: vegasanger.ac.uk
repeat_region	18739..18922	/rpt_family="AluY"		On Jan 20, 2002 this sequence version replaced gi:18250750.
repeat_region	complement(18923..19075)	/rpt_family="AluSx"		The following abbreviations are used to associate primary accession
repeat_region	complement(19076..19228)	/rpt_family="AluSx"		numbers given in the feature table with their source databases:
repeat_region	complement(19229..19381)	/rpt_family="AluSx"		Em, ENBL; Sw, SWISSPROT; Tr, TREMBL; Wp, WORMPEP; Information
repeat_region	complement(19382..19534)	/rpt_family="AluSx"		on the WORMPEP database can be found at
repeat_region	complement(19535..19687)	/rpt_family="AluSx"		http://www.sanger.ac.uk/Projects/Celegans/wormpep This sequence
repeat_region	complement(19688..19840)	/rpt_family="AluSx"		was generated from part of bacterial clone contigs of human
repeat_region	complement(19841..19993)	/rpt_family="AluSx"		chromosome 9, constructed by the Sanger Centre Chromosome 9 Mapping
repeat_region	complement(19994..20146)	/rpt_family="AluSx"		Group. Further information can be found at
repeat_region	complement(20147..20299)	/rpt_family="AluSx"		http://www.sanger.ac.uk/HGP/Chr9
repeat_region	complement(20300..20452)	/rpt_family="AluSx"		RP11-524G24 is from the library RP11-11.2 constructed by the group
repeat_region	complement(20453..20605)	/rpt_family="AluSx"		of Pieter de Jong. For further details see
repeat_region	complement(20606..20758)	/rpt_family="AluSx"		http://www.chori.org/bacpac/home.htm
repeat_region	complement(20759..20911)	/rpt_family="AluSx"		VECTOR: pBACe3.6
repeat_region	complement(20912..21064)	/rpt_family="AluSx"		----- Genome Center
repeat_region	complement(21065..21217)	/rpt_family="AluSx"		Center: Wellcome Trust Sanger Institute
repeat_region	complement(21218..21370)	/rpt_family="AluSx"		Web site: http://www.sanger.ac.uk
repeat_region	complement(21371..21523)	/rpt_family="AluSx"		Contact: vegasanger.ac.uk
repeat_region	complement(21524..21676)	/rpt_family="AluSx"		-----
repeat_region	complement(21677..21829)	/rpt_family="AluSx"		This sequence was finished as follows unless otherwise noted: all
repeat_region	complement(21830..21982)	/rpt_family="AluSx"		regions were either double-stranded or sequenced with an alternate
repeat_region	complement(21983..22135)	/rpt_family="AluSx"		chemistry or covered by high quality data (i.e., phred quality >=
repeat_region	complement(22136..22288)	/rpt_family="AluSx"		30); an attempt was made to resolve all sequencing problems, such
repeat_region	complement(22289..22441)	/rpt_family="AluSx"		as compressions and repeats; all regions were covered by at least
repeat_region	complement(22442..22594)	/rpt_family="AluSx"		one subclone; and the assembly was confirmed by restriction digest,
repeat_region	complement(22595..22747)	/rpt_family="AluSx"		except on the rare occasion of the clone being a YAC.

FEATURES

```

source
    Location/Qualifiers
    1..107529
    /organism="Homo sapiens"
    /mol_type="genomic DNA"
    /db_xref="taxon:9606"
    /chromosome="9"
    /clone="RP11-524G24"
    /clone_lib="RPCI-11.2"
    2000
    /note="Clone_right_end: RP11-162D16"
    8212..8295
    /note="Sequence from reads from a short insert library
    derived from a single pUC clone. Restriction digest data
    confirm the assembly. Sequence from overlapping clone
    RP11-508N19 (AL353609). Assembly confirmed by restriction
    digest."
    misc_feature
    9570..9633
    /note="Sequence from reads from a short insert library
    derived from a single pUC clone. Restriction digest data
    confirm the assembly. Sequence from overlapping clone
    RP11-508N19 (AL353609). Assembly confirmed by restriction
    digest."
    misc_feature
    9940..9982
    /note="Sequence from reads from a short insert library
    derived from a single pUC clone. Restriction digest data
    confirm the assembly. Sequence from overlapping clone
    RP11-508N19 (AL353609). Assembly confirmed by restriction
    digest."
    misc_feature
    9983..10001
    /note="Sequence from overlapping clone RP11-508N19
    (AL353609). Assembly confirmed by restriction digest."
    join(44539..44884,AL365274.17:50805..50908,
    AL365274.17:71452..71585,AL365274.17:129140..129293,
    AL365274.17:131025..131123,AL365274.17:132012..132075,
    AL365274.17:156611..157655)
    /gene="DAB2IP"
    /locus_tag="RP11-298A17.1-002"
    join(44539..44884,AL365274.17:50805..50908,
    AL365274.17:71452..71585,AL365274.17:129140..129293,
    AL365274.17:131025..131123,AL365274.17:132012..132075,
    AL365274.17:156611..157655)
    /gene="DAB2IP"
    /locus_tag="RP11-298A17.1-002"
    /product="DAB2 interacting protein"
    /note="match: ESTs: BF565256.1
    match: CDNAs: AK096391.1"
    join(44758..44884,AL365274.17:50805..50908,
    AL365274.17:71452..71585,AL365274.17:129140..129293,
    AL365274.17:131025..131123,AL365274.17:132012..132455)
    /gene="DAB2IP"
    /locus_tag="RP11-298A17.1-005"
    join(44758..44884,AL365274.17:50805..50908,
    AL365274.17:71452..71585,AL365274.17:129140..129293,
    AL365274.17:131025..131123,AL365274.17:132012..132455)
    /gene="DAB2IP"
    /locus_tag="RP11-298A17.1-005"
    /product="DAB2 interacting protein"
    /note="match: ESTs: BM545676.1 BM546615.1"
    join(44845..44884,AL365274.17:50805..50908,
    AL365274.17:71452..71585,AL365274.17:129140..129293,
    AL365274.17:131025..131123,AL365274.17:132012..132455)
    /gene="DAB2IP"
    /locus_tag="RP11-298A17.1-005"
    /standard_name="OTTHUMP00000022029"
    /codon_start=1
    /product="DAB2 interacting protein"
    /protein_id="CA115112.1"
    /db_xref="GI:55959205"
    /db_xref="InterPro:IPR001849"
    /db_xref="UniProt/TREMBL:Q5T4Q1"
    /translation="MEPDSLQDDSYESQPERFGRSRSLPGSLSEKSPSPMEPSAATP
    FRVTGFLKSGIKRTKSPQKLDNRHSFRHLPGFRSAAAAADNERSHLMPRLKE

```

```

SRSHESLLSPSSAVSALDLSMEEEVWIKPVHSSILGQDYCFEVTTSKSKCFSCRSAA
ERDKWMENLRVHPNKDNRKNSRVEHLKJWIEAKDIPAKKKYSLCELCLDVLVARTT
GKLTDNVFGEHFEFNLPLRTVTVHLVRETQKKKKRNSYLGSLVSPAASVAGR
QFVEKWPVVTNPCKGKGGPMIRIKARVQITILPMWYKEFAEHT"
complement(join(52547..53668,53731..53870))
/locus_tag="RP11-524G24.2-001"
complement(join(52547..53668,53731..53870))
/locus_tag="RP11-524G24.2-001"
/note="match: ESTs: BM715539.1"
75919..76068
/note="Sequence from overlapping clone RP11-508N19
(AL353609). Assembly confirmed by restriction digest."
join(75924..76282,AL365274.17:50805..50909)
/gene="DAB2IP"
/locus_tag="RP11-298A17.1-003"
join(75924..76282,AL365274.17:50805..50909)
/gene="DAB2IP"
/product="DAB2 interacting protein"
/note="match: ESTs: BI489865.1"
77213..77257
/note="Sequence from overlapping clone RP11-508N19
(AL353609). Assembly confirmed by restriction digest."
105530
/note="clone_left_end: RP11-298A17"
ORIGIN
Query Match 46.9%; Score 234.8; DB 8; Length 107529;
Best Local Similarity 70.1%; Pred. No. 1.5e-55;
Matches 352; Conservative 0; Mismatches 132; Indels 18; Gaps 2;
Qy 3 GGGGCTCACTCTGTCGCCAGGCTGGAGTGGGGGTGATCTCAGATCACTCAACCT 62
Db 10020 GAGTCTCGCTCTGTCGCCAGGCTGGAGTGGGGGTGATCTCAGATCACTCAACCT 10079
Qy 63 CCATCTCTGGGTCAAAATAATCTTTCCTCAGCTCTCAAGTAGCTGGAGCTTGTAG 122
Db 10080 CTGGCTCTGGGTTCATGCCATTCTCTGCTCAGCTCCCAAGTAGCTGGACT---- 10134
Qy 123 CTCTCAAGTAGCTGGGACACACACCACATGCCAGCTAAATTTTGTGTGTTTGTGTT 182
Db 10135 -----ACAGGCACCTGCCACAGGCTGGCTAAATTTTGTATTTTGTATTTTGT 10184
Qy 183 AGAGACAGGTTTTCACCATGTGGCCAGGCTGGGTGACCTCCCTTTTAGATTCTCCTCAT 242
Db 10185 AGAGACAGGTTTTCACCATGTGGGTAGGATGCTCGATCTCTGACCTTGCTCGC 10244
Qy 243 CCTGCTCTATCTTCTCCCTTTCT---AATGCAGTATCCAGTTTCTTCTTATATCACA 299
Db 10245 CCATCTCAGCTCCCAAGTGTGGGATTACAGTGTGAGCCACACACCGGCTCTTT 10304
Qy 300 ATTATATTCTTATTATTATTTAGACAGATCTTGCTTTGCGCAAGGCTGGAGTACAG 359
Db 10305 TTTCTTTTACTTTTTTTTTTTTGTAGATGGAGTCTTGCTCTGTCACCCAGGCTGGAGT 10364
Qy 360 TGTGTCATCTCGGCTCACTGCAAGCTCCACCTGCTGGGTTCAGCCATTCTCCCGCTC 419
Db 10365 TAGACATCTTGGCTCACTGCAATCTCTCTCTCTGAGTTCAACAAATTTCTCTGCTC 10424
Qy 420 AGCCTCCCCAGTAGCTGGGACTAAAGCGCTGCCACACCGCCCGCTAAATTTTTTGTGA 479
Db 10425 AGCCTTCGAGTAGCTGGGATTACAAGTAGTACCACAAATGCCCGCGAAGTTTTTGTGA 10484
Qy 480 TTTTAAATAAGACGGGTTTC 501
Db 10485 TTTTGTAGTAGACGGGTTTC 10506
RESULT 15
AL356115 LOCUS
DEFINITION Human DNA sequence from clone RP11-486022 on chromosome 10 Contains
the 3' end of a novel gene (K1A11128, FLJ14262, FLJ25809) , the 5'

```

end of a Siah-interacting protein (SIP) (calyculin binding protein) pseudogene, ribosomal protein S3A pseudogene 5 (RPS3AP5) and two CpG islands, complete sequence.

ACCESSION
VERSION
KEYWORDS
SOURCE
ORGANISM

AL356115.9 GI:9795038
HTG; calyculin; FLJ14262; FLJ25809; KIAA1128; RPS3A5;
Siah-interacting; SIP.
Homo sapiens (human)

REFERENCE
AUTHORS
TITLE
JOURNAL

1 (bases 1 to 150151)
Whitehead.S.
Direct Submission
Submitted (13-MAY-2005) Wellcome Trust Sanger Institute, Hinxton,
Cambridgeshire, CB10 1SA, UK. E-mail enquiries: vegas@sanger.ac.uk

Clone requests: clonesrequest@sanger.ac.uk

On Aug 11, 2000 this sequence version replaced gi:9713038.
The following abbreviations are used to associate primary accession numbers given in the feature table with their source databases:

Em.: EMBL; Sw.: SWISSPROT; Tr.: TREMBL; Wp.: WORMPEP; Information

on the WORMPEP database can be found at
http://www.sanger.ac.uk/Projects/C_elegans/wormpep This sequence was generated from part of bacterial clone contigs of human chromosome 10, constructed by the Sanger Centre Chromosome 10 Mapping Group. Further information can be found at
<http://www.sanger.ac.uk/HGP/Chr10>

----- Genome Center
Center: Wellcome Trust Sanger Institute

Center code: SC
Web site: <http://www.sanger.ac.uk>
Contact: vegas@sanger.ac.uk

This sequence was finished as follows unless otherwise noted: all regions were either double-stranded or sequenced with an alternate chemistry or covered by high quality data (i.e., phred quality >= 30); an attempt was made to resolve all sequencing problems, such as compressions and repeats; all regions were covered by at least one subclone; and the assembly was confirmed by restriction digest, except on the rare occasion of the clone being a VAC.
RP11-486022 is from the library RPCI-11.2 constructed by the group of Pieter de Jong. For further details see
<http://www.chori.org/bacpac/home.htm>

VECTOR: pBACe3.6.
Location/Qualifiers

source
1..150151
/organism="Homo sapiens"
/mol_type="genomic DNA"
/db_xref="taxon:9606"
/chromosome="10"
/clone="RP11-486022"
/clone_lib="RPCI-11.2"

misc_feature

1 /notes="Clone_left_end: RP11-486022"

misc_feature

53062
/notes="Clone left end: RP11-470018"
join(complement(AC074321.10:101526..>102204),
complement(AC074321.10:100180..100376),
complement(AC074321.10:56134..56224),
complement(AC074321.10:48102..48264),
complement(AC074321.10:35288..35483),
complement(AC074321.10:9979..10062),
complement(AC074321.10:3537..3623),53890..53979,
76190..76274,89764..94832)

/gene="KIAA1128"

/locus_tag="RP11-486022.1-002"
join(complement(AC074321.10:101526..>102204),
complement(AC074321.10:100180..100376),
complement(AC074321.10:56134..56224),
complement(AC074321.10:48102..48264),
complement(AC074321.10:35288..35483),
complement(AC074321.10:9979..10062),
complement(AC074321.10:3537..3623),53890..53979,
76190..76274,89764..94832)

mRNA

complement(AC074321.10:3537..3623),53890..53979,
76190..76274,89764..94832)
/gene="KIAA1128"
/locus_tag="RP11-486022.1-002"
/product="KIAA1128"
/notes="match: cDNAs: AB032954 AB032954.1 AK024324.1"
join(complement(AC074321.10:80053..80142),
complement(AC074321.10:56134..56224),
complement(AC074321.10:48102..48264),
complement(AC074321.10:9979..10062),
complement(AC074321.10:3537..3623),53890..53979,
89764..90720)
/gene="KIAA1128"
/locus_tag="RP11-486022.1-003"
join(complement(AC074321.10:80053..80142),
complement(AC074321.10:56134..56224),
complement(AC074321.10:48102..48264),
complement(AC074321.10:9979..10062),
complement(AC074321.10:3537..3623),53890..53979,
89764..90185)
/gene="KIAA1128"
/locus_tag="RP11-486022.1-007"
join(complement(AC074321.10:21690..21778),
complement(AC074321.10:9979..10062),
complement(AC074321.10:3537..3623),53890..53979,
89764..90185)
/gene="KIAA1128"
/locus_tag="RP11-486022.1-007"
/product="KIAA1128"
/notes="match: ESTs: BI461360.1 BI461611.1"
join(complement(AC074321.10:35288..35452),
complement(AC074321.10:9979..10062),
complement(AC074321.10:3537..3623),53890..53979,
76190..76274,83609..83648,89764..90017)
/gene="KIAA1128"
/locus_tag="RP11-486022.1-008"
join(complement(AC074321.10:35288..35452),
complement(AC074321.10:9979..10062),
complement(AC074321.10:3537..3623),53890..53979,
76190..76274,83609..83648,89764..90017)
/gene="KIAA1128"
/locus_tag="RP11-486022.1-008"
/product="KIAA1128"
/notes="match: ESTs: BQ722236.1"
join(complement(AC074321.10:21690..21778),
complement(AC074321.10:9979..10062),
complement(AC074321.10:3537..3623),
complement(AC074321.10:849..963),53890..53979,
89764..90011)
/gene="KIAA1128"
/locus_tag="RP11-486022.1-005"
join(complement(AC074321.10:21690..21778),
complement(AC074321.10:9979..10062),
complement(AC074321.10:3537..3623),
complement(AC074321.10:849..963),53890..53979,
89764..90011)
/gene="KIAA1128"
/locus_tag="RP11-486022.1-005"
/product="KIAA1128"
/notes="match: ESTs: BG719330.1 BQ429698.1"
join(complement(AC074321.10:48769..48813),
complement(AC074321.10:48102..48264),
complement(AC074321.10:35288..35483),
complement(AC074321.10:9979..10062),
complement(AC074321.10:3537..3623),53890..53979,
89764..90011)

```
89764..89976)
/locus_tag="RP11-486022.1-006"
/join(complement(AC074321.10:48769..48813),
complement(AC074321.10:48102..48264),
complement(AC074321.10:35288..35483),
complement(AC074321.10:9979..10062),
complement(AC074321.10:3537..3623),53890..53979,
89764..89976)
/gene="KIAA1128"
/locus_tag="RP11-486022.1-006"
/product="KIAA1128"
/note="match: ESTs: BQ40427.1"
Join(complement(AC074321.10:101526..102204),
complement(AC074321.10:100180..100376),
complement(AC074321.10:56134..56224),
complement(AC074321.10:48102..48264),
complement(AC074321.10:35288..35483),
complement(AC074321.10:9979..10062),
complement(AC074321.10:3537..3623),53890..53979,
76190..76274,89764..89858)
/gene="KIAA1128"
/locus_tag="RP11-486022.1-006"
/standard_name="OTHUMP0000020001"
/note="match: proteins: Q9D6G6 Q9H1N6 Q9H2S0 Q9H7U1
Q9ULU1"
/codon_start=1
/product="KIAA1128"
/protein_id="CAI40902.1"
/db_xref="GI:57209021"
/db_xref="UniProt/TREMBL:Q9ULU1"
/translation="AVDLTKYQNLQSLRVPVLPVSLKSRSPFSGTMTVDGNKNSPADTCV
NRPYAAGKKLALPNGPGVTGLGYRVMVHPSLLKSRSPFSGTMTVDGNKNSPADTCV
BEDATVLAKDRAANKDQELIENESVTKNQTMDKADKRYLSDDVDISSLSSSD
KNDLSEDSDFDIEDSNRTRITPEENSLKEEKHENGPPQDMFDSKREKAFSKTD
ENIDLSVDSRECTKHTSGNNLVSPDTRAGSFEPSDSSTGYMWDDEGLEPIG
NVHPVGSYESSEMSIDILNLESCDDLEDLDLVDLPEDAPLENVEDNNRFDPR
DNVRPQOGEQWKPQVQSGQEHYLSHPDHYHHKSDLSRGSPPRESPLGHFESY
GMPFQOQKMFVDVPTWVILDEMTLHMVQDCITAVTQLLKLRLHQRHDSGSLH
DIOLSPSSPEPDGKVYKNEIDLNEIKLKDEIKKLEQLQLATQHICHQK
CKEECTYADKYTOTPRRIRPGYSAPFSFQSGSFQGIPTVPPHRRQTSSTAFQ
FSQTHRSHPGKINAKATTYRGQP"
90701..90706
/gene="KIAA1128"
/locus_tag="RP11-486022.1-003"
90722
/gene="KIAA1128"
/locus_tag="RP11-486022.1-002"
91100..91105
/gene="KIAA1128"
/locus_tag="RP11-486022.1-002"

polyA_signal
/gene="KIAA1128"
/locus_tag="RP11-486022.1-003"
90722
/gene="KIAA1128"
/locus_tag="RP11-486022.1-002"
91100..91105
/gene="KIAA1128"
/locus_tag="RP11-486022.1-002"

Query Match 46.7%; Score 234.2; DB 8; Length 150151;
Best Local Similarity 72.9%; Pred. No. 2.3e-55;
Matches 364; Conservative 0; Mismatches 113; Indels 22; Gaps 4;

QY 3 GGGGCTCACTGTGCGCCAGGCTGGAGTGGGTGGGTGATCTCAGATCACTACACCT 62
DB 137980 GAGTCTGGCTCTGTAGCCAGGCTGGAGTGGGTGGGTGATCTCAGGCTACCAAGCT 138039

QY 63 CCATCTCTGGGTTCAATAATTCCTTCCTCAGCTCTCAAGTAGCTGGGACTTGTAG 122
DB 138040 CCACCTCTGTTTCACACCATCTCTGCTGCTCAGCTCCCGAGTAGCTGGGACT 138094

QY 123 CTCTCAAGTAGTGGGACACACCATGCGCCAGCTAATTTTTTGTGTGTTTTTGGT 182
DB 138095 -----ACAGGCGCCCGCCACACCGCCGGCTAATTTTT---TGTATTTTGTAGT 138140

QY 183 AGAGACAGTTTTCACCATGTTGGCAGGCTGGGTGACCTCCCTTTAGATTCCTCAT 242
DB 138141 GGAGACGGGGTTTACCGGTGTAGCAGGAT-GGTCTCGATCTCTGACCTCGTATCTG 138199

QY 243 CTTGCTCTATTCTTCCCTTTCTAATGCAATGATCCAGTTTCTCTTACTTATCACATTTATT 302
```

```
Db 138200 CTTGCTCAGCCT--CCAAAAGTGTCTGGGATTACAGGCGTGACTAAGGTTTCATTTTGTCT 138257
QY 303 ATTATTCTTATATTTATTTGACACAGAGTCTTTGCTTTTGTCCCAAGGCTGGAGTACAGTGG 362
DB 138258 GATTTTTTTTTTTTTTTTTTTTGTAGACAGAGTCTCGCTCTGTGCGCCAGGCTGGAGTGCAGCG 138317
QY 363 TGGCATCTCGGCTCACTGCAAGCTCCACCTGCTGGGTTTCAGGCCAATTCCTCGGCTCAGC 422
DB 138318 CGCCATCTCGGCTCACTGGAAGCTCCGCTCCCGGTTTCAGGCCAATTCCTCGCTTACG 138377
QY 423 CTCCTCAGTAGCTGGGACTAAAGCGCTGTGCAACACGCCCGCTTAATTTTTTGTATTT 482
DB 138378 CTCCTCGGTAGCTGGGAGTACAGGCGCCCGCCACACGCCCGGCTAATTTTTTGTATTT 138437
QY 483 TTAATAAGACGGGGTTTC 501
DB 138438 TTAGTAGAGACGGGTGTTTC 138456
```

Search completed: February 9, 2006, 18:20:26
Job time : 2701.09 secs

THIS PAGE BLANK (USPTO)

GenCore version 5.1.7
Copyright (c) 1993 - 2006 Bioacceleration Ltd.

OM nucleic - nucleic search, using sw model

Run on: February 9, 2006, 13:16:14 ; Search time 344.959 Seconds
(without alignments)
9679.442 Million cell updates/sec

Title: US-10-607-806-1_COPY_5000_5500

Perfect score: 501

Sequence: 1 acggggtcactctgtcgcc.....tttaataaagacggggttc 501

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 4996997 seqs, 3332346308 residues

Total number of hits satisfying chosen parameters: 9993994

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : N_Geneseq_21.*

- 1: Geneseqn1980s.*
- 2: Geneseqn1990s.*
- 3: Geneseqn2000s.*
- 4: Geneseqn2001as.*
- 5: Geneseqn2001bs.*
- 6: Geneseqn2002as.*
- 7: Geneseqn2002bs.*
- 8: Geneseqn2003as.*
- 9: Geneseqn2003bs.*
- 10: Geneseqn2003cs.*
- 11: Geneseqn2003ds.*
- 12: Geneseqn2004as.*
- 13: Geneseqn2004bs.*
- 14: Geneseqn2005s.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	501	100.0	12174	12 ADI35082	ADI35082 Human PLA
2	501	100.0	12174	12 ADJ09983	Adj09983 Human pho
3	501	100.0	13612	6 ABK47376	Abk47376 Human pho
4	429.2	85.7	3375	3 AAA35101	Aaa35101 Human ade
5	429.2	85.7	3375	3 AAF21223	Aaf21223 Human low
6	429.2	85.7	3375	10 ABZ96917	Abz96917 Human nuc
7	429.2	85.7	3375	11 ABD20766	Abd20766 Human pul
8	429.2	85.7	6771	3 AAA35103	Aaa35103 Human ade
9	429.2	85.7	6771	3 AAF21225	Aaf21225 Human low
10	429.2	85.7	6771	10 ABZ96919	Abz96919 Human nuc
11	429.2	85.7	6771	11 ABD20768	Abd20768 Human pul
c 12	231.6	46.2	35973	6 ABK13076	Abk13076 Human amy
c 13	231.4	46.2	602	4 AAL06681	Aal06681 Human rep
c 14	231.4	46.2	602	4 ABA08020	Aba08020 Human ova
c 15	228.8	45.7	127508	13 ABD33171	Abd33171 Human can
c 16	228.4	45.6	73100	14 ADZ42280	Adz42280 Human end
c 17	228.4	45.6	109906	6 ABK94411	Abk94411 DNA encod
c 18	228.4	45.6	109906	12 ADL08112	Adl08112 Human gen
c 19	227.6	45.4	10331	4 AAK79915	Aak79915 Human imm

c	20	227.6	45.4	10331	4 AAL03344	Aal03344 Human rep
c	21	224.2	44.8	129042	11 ACN44674	Acn44674 Human gen
c	22	224	44.7	14175	4 AAS27814	Aas27814 DNA encod
c	23	224	44.7	14175	4 AAK78859	Aak78859 Human imm
c	24	224	44.7	14175	10 ADB94617	Adb94617 Novel hum
c	25	223	44.5	260209	6 ABS56564	Abs56564 Human SUL
c	26	223	44.5	260209	12 ADN16204	Adn16204 Human sul
c	27	222.8	44.5	73995	11 ACN43986	Acn43986 Human gen
c	28	222.6	44.4	3067	10 ADE09851	Ade09851 Novel DNA
c	29	222.6	44.4	36534	11 ADL82795	Adl82795 Human sem
c	30	222.2	44.4	24167	5 ABA16132	Abal6132 Human ner
c	31	222	44.3	33769	5 ABA82622	Abas2622 Human HBM
c	32	222	44.3	33769	6 ABK22781	Abk22781 Human hig
c	33	222	44.3	33769	8 ACC45363	Acc45363 Human HBM
c	34	222	44.3	33769	10 ADB98063	Adb98063 HBM-relat
c	35	222	44.3	33769	10 ADE82432	Ade82432 Human DNA
c	36	222	44.3	33769	13 ADRI6926	Adri6926 BAC clone
c	37	222	44.3	33769	13 ADRA7577	Adra7577 BAC clone
c	38	222	44.3	33769	14 AEB69306	Aeb69306 Human hig
c	39	222	44.3	137046	13 ABD32936	Abd32936 Human can
c	40	222	44.3	156843	11 ACN44786	Acn44786 Human gen
c	41	221.8	44.3	12595	4 AAS42100	Aas42100 Genomic s
c	42	221	44.1	611	4 AAK71483	Aak71483 Human imm
c	43	220.4	44.0	890	4 AAH99337	Aah99337 Human pro
c	44	220.4	44.0	3884	4 AAH98627	Aah98627 Human EST
c	45	220.2	44.0	32152	4 ABA08132	Aba08132 Human ova

ALIGNMENTS

RESULT 1

ADI35082

ID ADI35082 standard; DNA; 12174 BP.

XX AC ADI35082;

XX DT 22-APR-2004 (first entry)

XX DE Human PLA2G1B nucleotide sequence.

XX KW PLA2G1B ; fat deposition; leanness; polymorphism;

XX KW non-insulin dependent diabetes mellitus; NIDDM; hyperinsulinemia;

XX KW hypertension; Glucose intolerance; dyslipidemia; hypercoagulability;

XX KW microalbuminuria; human; gene; ds.

XX OS Homo sapiens.

XX PN WO2004002295-A2.

XX PD 08-JAN-2004.

XX PF 27-JUN-2003; 2003WO-US020830.

XX PR 27-JUN-2002; 2002US-0392361P.

XX PA (SEQU-) SEQUENOM INC.

XX PI Adam GIR, Langdown ML;

XX DR WPI; 2004-082843/08.

XX DR P-PSDB; ADI35083.

XX PT Diagnosing a predisposition to fat deposition or leanness, useful for

XX PT diagnosing a predisposition to e.g. diabetes or hypertension, comprises

XX PT detecting the presence of a polymorphism in the PLA2G1B nucleic acid from

XX PT the subject.

XX PS Claim 1; SEQ ID NO 1; 91pp; English.

XX CC The invention relates to diagnosing a predisposition to fat deposition or

XX CC leanness in a subject comprising detecting the presence or absence of a

XX CC polymorphic variation associated with fat deposition at a polymorphic

CC presence or absence of these single nucleotide polymorphisms of PLA2G1B,
CC in particular G7328A and T9182G, and subsequently provide treatment that
CC reduces fat deposition. This treatment may consist of an appetite
CC suppressant, a lipase inhibitor, a phospholipase inhibitor, an exercise
CC regimen, a dietary regimen, psychological counselling, psychotherapy or a
CC psychotherapeutic. Accordingly, PLA2G1B is a target for reducing fat
CC deposition and it can be used to treat both obesity and non-insulin
CC dependent diabetes mellitus (NIDDM), as well as cardiovascular disorders
CC such as hypertension. As such, it exhibits anti-diabetic activity. This
CC polynucleotide sequence is the human PLA2G1B DNA of the invention.
XX

SQ Sequence 12174 BP; 3220 A; 2996 C; 2739 G; 3219 T; 0 U; 0 Other;

Query Match 100.0%; Score 501; DB 12; Length 12174;
Best Local Similarity 100.0%; Pred. No. 3.7e-133;
Matches 501; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 ACGGGGCTCCTCTGCGCCAGGCTGGAGTGGCGGTGCTGATCTCAGATCACTACAC 60
Db |||||||
QY 61 CTCATCTCCTGGGTTCAAAATAATCTCTGCTCAGGCTCTCAAGTAGCTGGGACTTGT 120
Db |||||||
QY 5060 CTCATCTCCTGGGTTCAAAATAATCTCTGCTCAGGCTCTCAAGTAGCTGGGACTTGT 5119
Db |||||||
QY 121 AGCTCTCAAGTAGCTGGGACACACACCATGCCCCAGCTAAATTTTGTGTGTTTTTTG 180
Db |||||||
QY 5120 AGCTCTCAAGTAGCTGGGACACACACCATGCCCCAGCTAAATTTTGTGTGTTTTTTG 5179
Db |||||||
QY 181 GTAGAGACAGGTTTTCACATGTTGGCGGAGTGGGCTGACCTCCCTTTTGTAGATCTCCTC 240
Db |||||||
QY 5180 GTAGAGACAGGTTTTCACATGTTGGCGGAGTGGGCTGACCTCCCTTTTGTAGATCTCCTC 5239
QY 241 ATCTGCTCTATCTTCCCTTTCTAATGCAATGCAATGCAATGCAATGCAATGCAATGCA 300
Db |||||||
QY 5240 ATCTGCTCTATCTTCCCTTTCTAATGCAATGCAATGCAATGCAATGCAATGCAATGCA 5299
Db |||||||
QY 301 TTATTATCTTATTATTATTATTATTATTATTATTATTATTATTATTATTATTATTATTATT 360
Db |||||||
QY 5300 TTATTATCTTATTATTATTATTATTATTATTATTATTATTATTATTATTATTATTATTATT 5359
QY 361 GGTGGATCTCGGCTCACTGCAAGCTCACCCTGCTGGGTTTCAAGGCTTCCCGGCTTCA 420
Db |||||||
QY 5360 GGTGGATCTCGGCTCACTGCAAGCTCACCCTGCTGGGTTTCAAGGCTTCCCGGCTTCA 5419
QY 421 GCCTCCCGGCTGAGTGGGCTTAAAGGCGCTGACACACGCGCTTAAAGGCTTAAAGGCTT 480
Db |||||||
QY 5420 GCCTCCCGGCTGAGTGGGCTTAAAGGCGCTGACACACGCGCTTAAAGGCTTAAAGGCTT 5479
QY 481 TTTTAAATAAGACGGGGTTTC 501
Db |||||||
QY 5480 TTTTAAATAAGACGGGGTTTC 5500

RESULT 3
ABK47376
ID ABK47376 standard; DNA; 13612 BP.

AC ABK47376;

DT 18-JUN-2002 (first entry)

DE Human Phospholipase A2, group1B (PLA2G1B) gene.

KW Human; ds; gene; SNP; single nucleotide polymorphism; pancreatitis;
KW pancreatic cancer; Phospholipase A2 group1B; PLA2G1B; gene therapy;
KW haplotype; genotype; chromosome 12q23-q24.1; transgenic; drug screening.

OS Homo sapiens.

XX Key Location/Qualifiers
FH variation replace (3845.A)
FT /*tag= a

FT FT /label= SNP
FT FT /note= "Single nucleotide polymorphism"
FT FT replace (3968.A)
FT FT /*tag= b
FT FT /label= SNP
FT FT /note= "Single nucleotide polymorphism"
FT FT 4053. .9613
FT FT /*tag= c
FT FT /product= "Pla2G1B"
FT FT 4053. .4086
FT FT /*tag= d
FT FT /number= 1
FT FT 4087. .5785
FT FT /*tag= e
FT FT /number= 1
FT FT 5786. .5945
FT FT /*tag= f
FT FT /number= 2
FT FT 5946. .6744
FT FT /*tag= g
FT FT /number= 2
FT FT replace (6060.A)
FT FT /*tag= h
FT FT /label= SNP
FT FT /note= "Single nucleotide polymorphism"
FT FT 6745. .6873
FT FT /*tag= i
FT FT /number= 3
FT FT replace (6844.A)
FT FT /*tag= j
FT FT /label= SNP
FT FT /note= "Single nucleotide polymorphism"
FT FT 6874. .9489
FT FT /*tag= k
FT FT /number= 3
FT FT 9490. .9613
FT FT /*tag= l
FT FT /number= 4
FT FT replace (9531.A)
FT FT /*tag= m
FT FT /label= SNP
FT FT /note= "Single nucleotide polymorphism"

WO200212562-A2.

14-FEB-2002.

06-AUG-2001; 2001WO-US024663.

04-AUG-2000; 2000US-0223179P.

(GENA-) GENAISSANCE PHARM INC.

Kazemi A, Kliem SE, Koshiy B;

WPI: 2002-303982/34.

P-PSDB; AAU78667.

Novel isolated human Phospholipase A2, Group 1B pancreas polynucleotide,
for therapeutic purposes, for studying expression and function of the
polynucleotide and for expressing the phospholipase protein.

Claim 1; Fig 1; Sipp; English.

CC The invention relates to an isolated human Phospholipase A2, Group 1B
CC (pancreas) (PLA2G1B) polynucleotide comprising a sequence which is a
CC polymorphic variant for a reference sequence for the PLA2G1B gene or its
CC fragment, or a polymorphic variant of a reference sequence for a PLA2G1B
CC cDNA or its fragment. Also included are haplotyping/genotyping the
CC PLA2G1B gene of an individual, predicting the haplotype pair for the
CC PLA2G1B gene of an individual, identifying an association between a trait
CC and at least one haplotype or haplotype pair of the PLA2G1B gene, an
CC isolated genotyping oligonucleotide for detecting a polymorphism in the

CC PLA2G1B gene, a recombinant non-human organism transformed or transfected
CC with the PLA2G1B sequence, where the organism expresses a PLA2G1B protein
CC encoded by the first nucleotide sequence or by the polymorphic variant
CC sequence, an isolated polypeptide comprising a sequence which is a
CC polymorphic variant of a reference sequence for the PLA2G1B protein or
CC its fragment, an anti-PLA2G1B monoclonal antibody, screening for drugs
CC targeting PLA2G1B, a computer system for storing and analysing
CC polymorphism data for the PLA2G1B gene and a genome anthology for PLA2G1B
CC gene. The PLA2G1B variant is useful in studying the expression and
CC function of PLA2G1B, and in expressing PLA2G1B protein for use in
CC screening for candidate drugs to treat diseases related to PLA2G1B
CC activity (e.g. pancreatitis and pancreatic cancer) and for therapeutic
CC purposes. The transgenic organism is useful for studying expression of
CC the PLA2G1B isogenes in vivo, for in vivo screening and testing of drugs
CC targeted against PLA2G1B protein, and for testing the efficacy of
CC therapeutic agents and compounds in a biological system. The antibody is
CC useful for studying the effect of the variation on the biological
CC activity of PLA2G1B as well as on the binding affinity of candidate drugs
CC targeting PLA2G1B. The present sequence is the PLA2G1B gene which is
CC located on chromosome 12q23-q24.1
XX
SQ Sequence 13612 BP; 3637 A; 3290 C; 3070 G; 3615 T; 0 U; 0 Other;

Query Match 100.0%; Score 501; DB 6; Length 13612;
Best Local Similarity 100.0%; Pred. No. 3.9e-133;
Matches 501; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 ACGGGGCTCAGCTCTGTCGCCAGGCTGGAGTGGCGTGGCTGATCTCAGATCAGTACAAAC 60
DB |||||||
4516 ACGGGGCTCAGCTCTGTCGCCAGGCTGGAGTGGCGTGGCTGATCTCAGATCAGTACAAAC 4575
QY 61 TCCCATCTCTCGGTTCAAAATTAATTTCTTTGGCTCAGCTCTCAAGTACGTGGGACTTGT 120
DB TCCCATCTCTCGGTTCAAAATTAATTTCTTTGGCTCAGCTCTCAAGTACGTGGGACTTGT 4635
QY 121 AGCTCTCAAGTAGCTGGGACACACACACATGCCAGCTAAATTTTGTGTTTTTTTGG 180
DB AGCTCTCAAGTAGCTGGGACACACACACATGCCAGCTAAATTTTGTGTTTTTTTGG 4695
QY 181 GTAGAGACAGGTTTTTCAACATGTTGGCCAGGCTGGGTGACCTCCCTTTTAGATTTCTCCTC 240
DB GTAGAGACAGGTTTTTCAACATGTTGGCCAGGCTGGGTGACCTCCCTTTTAGATTTCTCCTC 4755
QY 241 ATCTGCTCTATTCTTCCCTTTCTTAATGAGTATCCAGTTTCTTATCATTATCAATTTA 300
DB ATCTGCTCTATTCTTCCCTTTCTTAATGAGTATCCAGTTTCTTATCATTATCAATTTA 4815
QY 301 TTATTTATTTCTTATTATTATGACAGAGCTCTTGTCTGCGCAAGGCTGGAGTACAGT 360
DB TTATTTATTTCTTATTATTATGACAGAGCTCTTGTCTGCGCAAGGCTGGAGTACAGT 4875
QY 361 GGTGGGATCTCGGCTCAGCTGCAAGCTCCACCTGCTGGGTTTCAGCCATTCTCCCGCTCA 420
DB GGTGGGATCTCGGCTCAGCTGCAAGCTCCACCTGCTGGGTTTCAGCCATTCTCCCGCTCA 4935
QY 421 GCCTCCCGGAGTAGCTGGGACTAAAGCGGCTGCGCACCGCCCGCTTAATTTTGTAT 480
DB GCCTCCCGGAGTAGCTGGGACTAAAGCGGCTGCGCACCGCCCGCTTAATTTTGTAT 4995
QY 481 TTTTAAATAAGCGGGTTTC 501
DB TTTTAAATAAGCGGGTTTC 5016

RESULT 4

AAA35101

ID AAA35101 standard; DNA; 3375 BP.

XX AAA35101;

XX AAA35101;

XX 28-JUL-2000 (first entry)

XX Human adenosine receptor related polynucleotide SEQ ID NO:2790.

XX Human, adenosine receptor; low adenosine antisense oligonucleotide;
KW phosphorothioate; impaired respiration; inflammation; inhibitor; antiinflammatory;
KW allergic disease; bronchoconstriction; analgesic; impaired airway;
KW antiallergic; antiasthmatic; cytostatic; pulmonary vasoconstriction; asthma;
KW lung disease; ischaemic condition; pain; cystic fibrosis; emphysema;
KW respiratory distress syndrome; chronic obstructive pulmonary disease; COPD;
KW pulmonary hypertension; lymphoma; carcinoma; metastasis; ss.
XX Homo sapiens.
OS
XX WO200009525-A2.
PN
XX 24-FEB-2000.
FD
XX 03-AUG-1999; 99WO-US017712.
PF
XX 03-AUG-1998; 98US-0095212P.
PR
XX (UYEC-) UNIV EAST CAROLINA.
PA
XX Nyce JW;
XX WPI; 2000-205971/18.
XX
XX New antisense oligonucleotides useful for treating e.g. pulmonary
PT vasoconstriction, inflammation, allergies, asthma, hypertension,
PT bronchitis, emphysema, respiratory distress syndrome, ischemia or
PT cancers.
XX
PS Disclosure; Page 1056-1057; 1343pp; English.

CC The present invention describes a new composition comprising an antisense
CC oligonucleotide (ON) with low adenosine (up to 15%), which targets
CC nucleic acids involved in bronchoconstriction, allergies, and/or
CC inflammation. The ON can have antiinflammatory, antiallergic,
CC antiasthmatic, cytostatic and analgesic activities. The compositions are
CC useful for the treatment of diseases associated with inflammation,
CC impaired airways, including lung disease and diseases whose secondary
CC effects afflict the lungs of a subject. They can be used for treating
CC e.g. ischaemic conditions, pulmonary vasoconstriction, allergies, asthma,
CC impaired respiration, respiratory distress syndrome, pain, cystic
CC fibrosis, pulmonary hypertension, emphysema, chronic obstructive
CC pulmonary disease (COPD), and cancers such as leukaemias, lymphomas,
CC carcinomas, and cancers which may metastasise to the lungs, including
CC breast and prostate cancer. The reduction of the adenosine content of the
CC ONs reduces side effects. The A-containing ONs break down with the
CC release of deoxyadenosine which activates adenosine receptors causing
CC bronchoconstriction and inflammation. AAA32313 to AAA35312 represent the
CC nucleotide sequences given in the sequence listing from the present
CC invention, which correspond to SEQ ID NO:1 to 2815, and then the last 185
CC sequences are also called SEQ ID NO:1 to 185, but the sequences differ
CC from the previously named sequences. SEQ ID NO:11 to 1680 (AAA32323 to
CC AAA33992) are specifically claimed ONs from the present invention. N.B.
CC Sequences given in the disclosure of the present invention do not match
CC up with their corresponding SEQ ID NO: sequences given in the sequence
CC listing

SQ Sequence 3375 BP; 837 A; 866 C; 764 G; 908 T; 0 U; 0 Other;
Query Match 85.7%; Score 429.2; DB 3; Length 3375;
Best Local Similarity 97.4%; Pred. No. 9.8e-113;
Matches 489; Conservative 0; Mismatches 8; Indels 5; Gaps 5;

QY 1 ACGGGGCTCAGCTCTGTCGCCAGGCTGGAGTGGCGTGGCTGATCTCAGATCAGTACAAAC 60
DB |||||||
996 ACGGGGCTCAGCTCTGTCG-CCAGGCTGGAGTGGGTGGGCTGATCTCAGATCAGTACAAAC 1054
QY 61 CTCCCATCTCCTCGGTTCAAAATTAATTTCTTTGGCTCAGCTCTCAGTACGACTTGT 120
DB |||||||
1055 CTCCCATCTCCT-GGTTCAAAATTAATTTCTTTGGCTCAGCTCTCAGTACGACTTGT 1112


```

RESULT 6
ABZ96917
ID ABZ96917 standard; DNA; 3375 BP.
XX
AC ABZ96917;
XX
DT 17-OCT-2003 (first entry)
DE Human nucleic acid sequence.
XX
KW Human; antisense; lung dysfunction; nasal airway dysfunction;
KW antiinflammatory steroid; ubiquinone; antiinflammatory; antiallergic;
KW antiaschmatic; hypotensive; immunosuppressive; cytostatic; gene therapy;
KW antisense gene therapy; respiratory; lung; adenosine sensitivity;
KW adenosine receptor; bronchodilation; bronchoconstriction; lung allergy;
KW lung inflammation; respiratory disease; ds.
XX
OS Homo sapiens.
XX
XX WO200285308-A2.
XX
XX 31-OCT-2002.
XX
XX 23-APR-2002; 2002WO-US013135.
XX
XX 24-APR-2001; 2001US-0286137P.
XX
XX (EPIC-) EPIGENESIS PHARM INC.
XX
XX Nyce JW, Li Y, Sandrasagra A, Katz E, Pabalan J, Aguilar D;
XX Miller S, Tang L, Shahabuddin S;
XX WPI; 2003-229219/22.
XX
XX Pharmacological composition for treating ailments associated with impaired
XX respiration, has oligo(s) antisense to specific gene(s) or its
XX corresponding RNAs, and glucocorticoid or non-glucocorticoid steroid or
XX ubiquinone.
XX
XX Disclosure; SEQ ID NO 12159; 872pp; English.
XX
XX The invention relates to a novel pharmaceutical composition, which has a
XX first active agent comprising an oligonucleotide antisense to the
XX initiation codon, coding region, 5' or 3' end genomic flanking regions,
XX 5' and 3' intron-exon junctions, or regions within 2-10 nucleotides of
XX junctions of genes encoding a polypeptide associated with lung and/or
XX nasal airway dysfunction and a second active agent comprising an
XX antiinflammatory steroid and ubiquinone. A composition of the invention
XX has antiinflammatory, antiallergic, antiaschmatic, hypotensive,
XX immunosuppressive, and cytostatic activity. The composition may have a
XX use in antisense gene therapy. The composition is useful for treating or
XX preventing a respiratory, lung or malignant disease or condition, also
XX for enhancing the prophylactic or therapeutic respiratory effect of an
XX antiinflammatory steroid in a subject, for reducing or depleting levels
XX of, or reducing sensitivity to adenosine, reducing levels of ubiquinone or
XX receptor, producing bronchodilation, increasing levels of ubiquinone or
XX lung surfactant in a subject's tissue, or treating bronchoconstriction,
XX lung inflammation, lung allergies, or a respiratory disease or condition.
XX Note: The sequence data for this patent is not represented in the printed
XX specification, but was obtained in electronic format directly from WIPO
XX at ftp.wipo.int/pub/published_pct_sequences
XX
XX Sequence 3375 BP; 837 A; 866 C; 764 G; 908 T; 0 U; 0 Other;
XX
Query Match 85.78; Score 429.2; DB 10; Length 3375;
Best Local Similarity 97.4%; Pred. No. 9.8e-113;
Matches 489; Conservative 0; Mismatches 8; Indels 5; Gaps 5;
XX
XX 1 ACGGGGCTCACTCTGTGCGCCAGGCTGGAGTGGCGGTGATCTCAGATCACTACAAC 60
XX
XX 996 ACGGGGCTCACTCTGTGCG-CCAGGCTGGAGTGGCGGTGATCTCAGATCACTACAAC 1054
XX
XX 61 CTCATCTCTGGGTTCAATAATTCCTTGGCTCAGCCTCTCAAGTAGCTGGACTTGT 120

```

```

Db 1055 CTCATCTCT-GGTCAATAATTCCTTGGCTCAGCCTCTCAAGTAGCT-GGACTTGT 1112
Qy 121 AGCTTCAAGTAGCTGGCAGACACACCACCATGCCAGTAATTTTCTGCTGTTTTTTTG 180
Db 1113 AGCTTCAAGTAGCTGGCAGACACACCACCATGCCAGTAATTTTCTGCTGTTTTTTTG 1172
Qy 181 GTAGAGACAGTTTTCACCATGTTGGCCAGGCTGGGTGACCTCCCTTTTAGATTTCTCCTC 240
Db 1173 GTAGAGACAGTTTTCACCATGTTGGCCAGGCTGGGTGACCTCCCTTTTAGATTTCTCCTC 1232
Qy 241 ATCTGCTCTATTTCTCCCTTTCT-AATGACAGTATCCAGTTTCTTACTTATCACATTT 299
Db 1233 ATCTGCTCTATTTCTCCCTTTCTAATGACAGTATCCAGTTTCTTACTTATCACATTT 1292
Qy 300 ATTAATTTCTTATTTATTTAGACAGAGTCTTGTCTGCGCAAGGCTGGAGTACAG 359
Db 1293 ATTATTATTTCTTATTTATTTAGACAGAGTCTTGTCTTGTGCCCAAGGCTGGAGTACAG 1352
Qy 360 TGGTGGATCTCGGCTCACTGCAAGCTCCACTCTGGGTTTCAGCCATTCTCCGCGCTC 419
Db 1353 TGGTGGATCTCGGCTCACTGCAAGCTCCACTCTGGGTTTCAGCCATTCTCCGCGCTC 1412
Qy 420 AGCTTCCCTCAGTAGCTGGGACTAAAGCGCTGCCACCCAGCCCGCTTAAATTTTGTGA 479
Db 1413 AGCTTCCCTCAGTAGCTGGGACTAAA-GGCGCTGCCACCCAGCCCGCTTAAATTTTGTGA 1471
Qy 480 TTTTATAAAGACGGGTTTC 501
Db 1472 TTTTATAAAGACGGGTTTC 1493
XX
RESULT 7
ABD20766
ID ABD20766 standard; DNA; 3375 BP.
XX
AC ABD20766;
XX
XX 29-JUL-2004 (first entry)
XX
XX Human pulmonary and inflammatory target DNA #377.
XX
XX Human; antisense; bronchoconstriction; allergy; hyposecretion; pain;
XX respiratory tract inflammation; adenosine sensitivity; lung; cancer;
XX surfactant depletion; antiallergic; antiinflammatory; antiaschmatic;
XX analgesic; hypotensive; immunosuppressive; cytostatic; cystic fibrosis;
XX beta-adrenergic agonist; respiratory disease; pulmonary vasoconstriction;
XX respiratory distress syndrome; allergic rhinitis; pulmonary hypertension;
XX emphysema; chronic obstructive pulmonary disease; cancer; bronchitis;
XX pulmonary transplantation rejection; ds.
XX
XX Homo sapiens.
XX
XX WO200285309-A2.
XX
XX 31-OCT-2002.
XX
XX 23-APR-2002; 2002WO-US013143.
XX
XX 24-APR-2001; 2001US-0286036P.
XX
XX (EPIC-) EPIGENESIS PHARM INC.
XX
XX Nyce JW, Li Y, Sandrasagra A, Katz E, Pabalan J, Aguilar D;
XX Miller S, Tang L, Shahabuddin S;
XX WPI; 2003-093058/08.
XX
XX Pharmaceutical composition for treating asthma, has antisense
XX oligonucleotide containing less percentage of adenosine, targeted to
XX nucleic acids associated with lung airway or lung dysfunction, and
XX bronchodilating agent.
XX

```

PS Claim 15; SEQ ID NO 12159; 763pp; English.

CC This invention describes a novel composition (a) a first active agent, comprising oligonucleotides, effective for alleviating

CC bronchoconstriction, respiratory tract inflammation, allergies and

CC reducing adenosine sensitivity, levels of adenosine (A) or (A) receptors, surfactant depletion or hyposecretion, when administered to a mammal. The

CC oligonucleotides are derived from a gene encoding or regulating

CC expression of a target polypeptide associated with lung airway or lung

CC dysfunction or cancer and can be anti-sense to the corresponding mRNA.

CC The invention also describes a kit, that comprises: (a) a delivery

CC device, in separate containers, (b) the oligonucleotides, (c)

CC instructions for adding a carrier and for use of the kit. The composition

CC of the invention has anti-allergic, anti-inflammatory, antiasthmatic,

CC analgesic, hypotensive, immunosuppressive and cytostatic activity, is a

CC beta-adrenergic agonist. The composition is useful for preventing or

CC treating a respiratory, lung or malignant disease. The administered

CC composition comprises oligo and is administered to reduce the production

CC or availability, or to increase the degradation of the target mRNA or to

CC reduce the amount of target polypeptide present in the lungs. The

CC pulmonary obstruction, and/or bronchoconstriction and/or lung

CC inflammation, allergies and/or surfactant hypoproduction are associated

CC with a disease or condition such as pulmonary vasoconstriction,

CC inflammation, allergies, asthma, impeded respiration, respiratory

CC distress syndrome, pain, cystic fibrosis, allergic rhinitis, pulmonary

CC hypertension, emphysema, chronic obstructive pulmonary disease, pulmonary

CC transplantation rejection, pulmonary infections, bronchitis or cancer.

CC The reduced adenosine content of the anti-sense oligos corresponding to

CC thymidines present in the target RNA serves to prevent the breakdown of

CC the oligonucleotides into products that free adenosine into the system

CC e.g., lung, brain, heart, kidney, etc, tissue environment and thereby, to

CC prevent any unwanted effects due to it

XX

Seq 3375 BP; 837 A; 866 C; 764 G; 908 T; 0 U; 0 Other;

Query Match 85.7%; Score 429.2; DB 11; Length 3375;

Best Local Similarity 97.4%; Pred. No. 9.8e-113;

Matches 489; Conservative 0; Mismatches 8; Indels 5; Gaps 5;

QY 1 ACGGGGCTCCTCTCTCGCCAGGCTGGAGTGGCGTGGCTGATCTCAGATCAGCTACAC 60

DB 996 ACGGGGCTCCTCTCTCTCG - CAGGCTGGAGTGGCGTGGCTGATCTCAGATCAGCTACAC 1054

QY 61 CTCCTCTCTGGGTTCAATAAATTTCTCTGCTCAGGCTCTCAAGTAGCTGGGACTTGT 120

DB 1055 CTCCTCTCTCT - GGTTCATAAATTTCTCTGCTCAGGCTCTCAAGTAGCT - GGACTTGT 1112

QY 121 AGCTTCTCAAGTAGCTGGGACACACACACATGCCAGCTAAATTTTGTGTTTTTTTGT 180

DB 1113 AGCTTCTCAAGTAGCTGGGACACACACACATGCCAGCTAAATTTTGTGTTTTTTTGT 1172

QY 181 GTAGAGACAGGTTTTTCCACATGTTGGCCAGGCTGGTGACCTCCCTTTTATGATCTCCTC 240

DB 1173 GTAGAGACAGGTTTTTCCACATGTTGGCCAGGCTGGTGACCTCCCTTTTATGATCTCCTC 1232

QY 241 ATCTGCTCTATTTCTTCCCTTTCT - AATGCAAGTATCAGATTTCTTACTTATCAGATTT 299

DB 1233 ATCTGCTCTATTTCTTCCCTTTCTAATGCAAGTATCAGATTTCTTACTTATCAGATTT 1292

QY 300 ATTATATTTCTTATTTATTTAGACAGAGCTTCTGTTTGTGGCAAGGCTGGAGTACAG 359

DB 1293 ATTATATTTCTTATTTATTTAGACAGAGCTTCTGTTTGTGGCAAGGCTGGAGTACAG 1352

QY 360 TGGTGGATCTCGGCTCACTGCAGCTCCACTGCTGGTGTTCAGCCATTTCTCCGCTC 419

DB 1353 TGGTGGATCTCGGCTCACTGCAGCTCCACTGCTGGTGTTCAGCCATTTCTCCGCTC 1412

QY 420 AGCTTCTCCAGTAGCTGGGACTAAAGGGGCTTGCACACACCCGCTGATTTTGTGTA 479

DB 1413 AGCTTCTCCAGTAGCTGGGACTTAA - GGCTTGCACACACCCGCTGATTTTGTGTA 1471

QY 480 TTTTATAAAGACGGGTTTC 501

Db 1472 TTTTATAAAGACGGGTTTC 1493

RESULT 8

AAA35103

ID AAA35103 standard; DNA; 6771 BP.

XX

AC AAA35103;

XX

DT 28-JUL-2000 (first entry)

XX

DE Human adenosine receptor related polynucleotide SEQ ID NO:2792.

XX

KW Human; adenosine receptor; low adenosine antisense oligonucleotide; phosphorothioate; impaired respiration; inflammation; allergy; allergic disease; bronchoconstriction; inhibitor; antiinflammatory; antiasthmatic; antiasthmatic; cytotstatic; analgesic; impaired airway; lung disease; ischaemic condition; pulmonary vasoconstriction; asthma; respiratory distress syndrome; pain; cystic fibrosis; emphysema; pulmonary hypertension; chronic obstructive pulmonary disease; COPD; cancer; leukaemia; lymphoma; carcinoma; metastasis; ss.

XX

OS Homo sapiens.

XX

FN WO200009525-A2.

XX

PD 24-FEB-2000.

XX

PF 03-AUG-1999; 99WO-US017712.

XX

PR 03-AUG-1998; 98US-0095212P.

XX

PA (UYEC-) UNIV EAST CAROLINA.

XX

PI Nyce JW;

XX

DR WPI; 2000-205971/18.

XX

PT New antisense oligonucleotides useful for treating e.g. pulmonary vasoconstriction, inflammation, allergies, asthma, hypertension, bronchitis, emphysema, respiratory distress syndrome, ischemia or cancers.

XX

PS Disclosure; Page 1058-1059; 1343pp; English.

XX

CC The present invention describes a new composition comprising an antisense oligonucleotide (ON) with low adenosine (up to 15%), which targets nucleic acids involved in bronchoconstriction, allergies, and/or inflammation. The ON can have antiinflammatory, antiallergic, antiasthmatic, cytotstatic and analgesic activities. The compositions are useful for the treatment of diseases associated with inflammation,

CC impaired airways, including lung disease and diseases whose secondary effects afflict the lungs of a subject. They can be used for treating

CC e.g. ischaemic conditions, pulmonary vasoconstriction, allergies, asthma,

CC impeded respiration, respiratory distress syndrome, pain, cystic

CC fibrosis, pulmonary hypertension, emphysema, chronic obstructive

CC pulmonary disease (COPD), and cancers such as leukaemias, lymphomas,

CC carcinomas, and cancers which may metastasise to the lungs, including

CC breast and prostate cancer. The reduction of the adenosine content of

CC ONs reduces side effects. The A-containing ONs break down with the

CC release of deoxyadenosine which activates adenosine receptors causing

CC bronchoconstriction and inflammation. AAA32313 to AAA35312 represent the

CC nucleotide sequences given in the sequence listing from the present

CC invention, which correspond to SEQ ID NO:1 to 2815, and then the last 185

CC sequences are also called SEQ ID NO:1 to 185, but the sequences differ

CC from the previously named sequences. SEQ ID NO:11 to 1680 (AAA32323 to

CC AAA33992) are specifically claimed ONs from the present invention. N.B.

CC Sequences given in the disclosure of the present invention do not match

CC up with their corresponding SEQ ID NO: sequences given in the sequence

CC listing

XX

Seq Sequence 6771 BP; 1815 A; 1570 C; 1439 G; 1947 T; 0 U; 0 Other;

Query Match 85.7%; Score 429.2; DB 3; Length 6771;
Best Local Similarity 97.4%; Pred. No. 1.3e-112;
Matches 489; Conservative 0; Mismatches 8; Indels 5; Gaps 5;

QY 1 ACGGGGCTCACTCTCTCGCCAGGCTGGAGTGGGTGGCTGATCTCAGATCACTACAAC 60
DB 1558 ACGGGGCTCACTCTCTCG-CCAGGCTGGAGTGGGTGGCTGATCTCAGATCACTACAAC 1616

QY 61 CTCATCTCTCGGGTTCAAATAATTTCTTGCTCAGCCTCTCAAGTAGCTGGAGCTTGT 120
DB 1617 CTCATCTCCT-GGTTCAAATAATTTCTTGCTCAGCCTCTCAAGTAGCT-GGACTTGT 1674

QY 121 AGCTCTCAAGTAGCTGGACACACACATGCCAGCTAAATTTTGTGTTTGTGTTTGTG 180
DB 1675 AGCTCTCAAGTAGCTGGACACACACATGCCAGCTAAATTTTGTGTTTGTGTTTGTG 1734

QY 181 GTAGAGACAGGTTTTTCAACATGTTGGCAGGCTGGGTGAGCTTCCCTTTTAGATTCTCCTC 240
DB 1735 GTAGAGACAGGTTTTTCAACATGTTGGCAGGCTGGGTGAGCTTCCCTTTTAGATTCTCCTC 1794

QY 241 ATCTGCTCTATTTCTCCCTTTCT-AATGCAGTATCCAGTTTCTTACTTATCACTTT 299
DB 1795 ATCTGCTCTATTTCTCCCTTTCTAAATGCAGTATCCAGTTTCTTACTTATCACTTT 1854

QY 300 ATTATTTCTTATTATTATTCAGACAGAGTCTTCTTGTGTCGCAAGCTGGAGTACAG 359
DB 1855 ATTATTTCTTATTATTATTCAGACAGAGTCTTCTTGTGTCGCAAGCTGGAGTACAG 1914

QY 360 TGGTGGATCTCGGCTCACTGCAAGCTCCACCTGCTGGGTTCCAGCCATCTCCGCTC 419
DB 1915 TGGTGGATCTCGGCTCACTGCAAGCTCCACCTGCTGGGTTCCAGCCATCTCCGCTC 1974

QY 420 AGCTCCCGCAGTAGCTGGAGCTAAAGGCGCTGCCACACGCCCCGCTAAATTTTGTGTA 479
DB 1975 AGCTCCCGCAGTAGCTGGAGCTAAAGGCGCTGCCACACGCCCCGCTAAATTTTGTGTA 2033

QY 480 TTTTAAATAAGACGGGGTTTC 501
DB 2034 TTTTAAATAAGACGGGGTTTC 2055

RESULT 9
AAF21225
ID AAF21225 standard; DNA; 6771 BP.
XX
AC AAF21225;
XT
DT 14-MAR-2001 (first entry)
DE Human low adenosine antisense oligonucleotide related sequence #2792.
XX
KW Low adenosine antisense oligonucleotide; phosphorothioate; allergy;
KW human; airway disorder; bronchoconstriction; lung inflammation;
KW surfactant depletion; respiratory; bronchodilator; antiinflammatory;
KW immunosuppressive; antiasthmatic; analgesic; hypotensive; cystostatic;
KW respiratory obstruction; pulmonary obstruction; impeded respiration;
KW surfactant hypoproduction; pulmonary vasoconstriction; asthma; RDS;
KW respiratory distress syndrome; pain; cystic fibrosis; allergic rhinitis;
KW pulmonary hypertension; emphysema; pulmonary transplantation rejection;
KW chronic obstructive pulmonary disease; pulmonary infection; bronchitis;
KW cancer; ss.
XX
OS Homo sapiens.
XX
PN W0200062736-A2.
XX
PD 26-OCT-2000.
XX
PF 24-MAR-2000; 2000WO-US008020.
XX
PR 06-APR-1999; 99US-0127958P.
XX
PA (UYEC-) UNIV EAST CAROLINA.

PA (NYCE/) NYCE J W.
XX Nyce JW;
XX WPI; 2000-679539/66.
XX Low adenosine (A) content antisense oligonucleotides which do not trigger
PT adenosine receptors during metabolism, useful e.g. for treating cancers
PT and respiratory obstructions.
PS Disclosure; Page 1137-1139; 1592pp; English.
XX The present invention describes low adenosine (A) content antisense
CC oligonucleotides and compositions (I) comprising them. In the antisense
CC oligonucleotides the A is replaced by a 'universal' or alternative base.
CC (I) can have respiratory, bronchodilator, antiinflammatory, analgesic,
CC immunosuppressive, antiasthmatic, hypotensive and cytostatic activities.
CC The antisense oligonucleotides and (I) can be used to down-regulate the
CC expression and/or activity of target polypeptides associated with
CC lung/respiratory disorders and malignancies, such as stimulating and
CC activating peptide factors and transmitters, transcription factors,
CC immunoglobulins and antibodies, antibody receptors, cytokines and
CC chemokines, endogenously produced specific and non-specific enzymes,
CC binding proteins, adhesion molecules and their receptors, cytokine and
CC chemokine receptors, adenosine receptors, bradykinin receptors, central
CC nervous system (CNS) and peripheral nervous and non-nervous system
CC receptors, CNS and peripheral nervous and non-nervous system peptide
CC transmitters, defensins, growth factors, vasoactive peptides and
CC receptors, binding proteins and malignancy associated proteins. The
CC antisense oligonucleotides may be used in this way to treat disorders
CC including respiratory obstruction (especially pulmonary obstruction
CC and/or bronchoconstriction) and/or lung inflammation, allergy(ies) and/or
CC surfactant hypoproduction which are associated with a disease or
CC condition selected from pulmonary vasoconstriction, inflammation,
CC allergies, asthma, impeded respiration, respiratory distress syndrome
CC (RDS), pain, cystic fibrosis (CF), allergic rhinitis (AR), pulmonary
CC hypertension, emphysema, chronic obstructive pulmonary disease (COPD),
CC pulmonary transplantation rejection, pulmonary infections, bronchitis,
CC and/or cancer. AAF18434 to AAF21543 represent human polynucleotide
CC fragments and antisense oligonucleotides used in the exemplification of
CC the present invention
XX
SQ Sequence 6771 BP; 1815 A; 1570 C; 1439 G; 1947 T; 0 U; 0 Other;

Query Match 85.7%; Score 429.2; DB 3; Length 6771;
Best Local Similarity 97.4%; Pred. No. 1.3e-112;
Matches 489; Conservative 0; Mismatches 8; Indels 5; Gaps 5;

QY 1 ACGGGGCTCACTCTCTCGCCAGGCTGGAGTGGGTGGCTGATCTCAGATCACTACAAC 60
DB 1558 ACGGGGCTCACTCTCTCG-CCAGGCTGGAGTGGGTGGCTGATCTCAGATCACTACAAC 1616

QY 61 CTCATCTCTCGGGTTCAAATAATTTCTTGCTCAGCCTCTCAAGTAGCTGGAGCTTGT 120
DB 1617 CTCATCTCCT-GGTTCAAATAATTTCTTGCTCAGCCTCTCAAGTAGCT-GGACTTGT 1674

QY 121 AGCTCTCAAGTAGCTGGACACACACATGCCAGCTAAATTTTGTGTTTGTGTTTGTG 180
DB 1675 AGCTCTCAAGTAGCTGGACACACACATGCCAGCTAAATTTTGTGTTTGTGTTTGTG 1734

QY 181 GTAGAGACAGGTTTTTCAACATGTTGGCAGGCTGGGTGAGCTTCCCTTTTAGATTCTCCTC 240
DB 1735 GTAGAGACAGGTTTTTCAACATGTTGGCAGGCTGGGTGAGCTTCCCTTTTAGATTCTCCTC 1794

QY 241 ATCTGCTCTATTTCTCCCTTTCT-AATGCAGTATCCAGTTTCTTACTTATCACTTT 299
DB 1795 ATCTGCTCTATTTCTCCCTTTCTAAATGCAGTATCCAGTTTCTTACTTATCACTTT 1854

QY 300 ATTATTTCTTATTATTATTCAGACAGAGTCTTCTTGTGTCGCAAGCTGGAGTACAG 359
DB 1855 ATTATTTCTTATTATTATTCAGACAGAGTCTTCTTGTGTCGCAAGCTGGAGTACAG 1914

QY 360 TGGTGGATCTCGGCTCACTGCAAGCTCCACCTGCTGGGTTCCAGCCATCTCCGCTC 419

Db 1915 TGGTGGAGTCTCGGCTCACTGAAGCTCCACCTGCTGGGTTACGCCATTCTCCGGCTC 1974
Qy 420 AGCTCTCCCGAGTAGCTGGGACTAAAGGGGCGCTGCGCACACCGCCCGCTAAATTTTGTGA 479
Db 1975 AGCTCTCCCGAGTAGCTGGGACTAAA-GCGCTGCGCACACCGCCCGCTAAATTTTGTGA 2033
Qy 480 TTTTAAATAAGACGGGGTTTC 501
Db 2034 TTTTAAATAAGACGGGGTTTC 2055

RESULT 10
ID ABZ96919 standard; DNA; 6771 BP.
XX AC ABZ96919;
XX DT 17-OCT-2003 (first entry)
XX DE Human nucleic acid sequence.
XX KW Human; antisense; lung dysfunction; nasal airway dysfunction;
KW antiinflammatory steroid; ubiqunone; antiinflammatory; antiallergic;
KW antiasthmatic; hypotensive; immunosuppressive; cyostatic; gene therapy;
KW antisense gene therapy; respiratory; lung; adenosine sensitivity;
KW adenosine receptor; bronchodilation; bronchoconstriction; lung allergy;
KW lung inflammation; respiratory disease; ds.
OS Homo sapiens.
XX WO200285308-A2.
XX 31-OCT-2002.
XX 23-APR-2002; 2002WO-US013135.
XX 24-APR-2001; 2001US-0286137P.
XX (EPTG-) EPIGENESIS PHARM INC.
XX Nyce JW, Li Y, Sandrasagra A, Katz E, Pabalan J, Aguilar D;
XX Miller S, Tang L, Shahabuddin S;
XX WPI; 2003-229219/22.
XX Pharmaceutical composition for treating ailments associated with impaired
XX respiration, has oligo(s) antisense to specific gene(s) or its
XX corresponding RNAs, and glucocorticoid or non-glucocorticoid steroid or
XX ubiqunone.
XX Disclosure; SEQ ID NO 12161; 872pp; English.
XX The invention relates to a novel pharmaceutical composition, which has a
XX first active agent comprising an oligonucleotide antisense to the
XX initiation codon, coding region, 5' or 3' end genomic flanking regions,
XX 5' and 3' intron-exon junctions, or regions within 2-10 nucleotides of
XX junctions of genes encoding a polypeptide associated with lung and/or
XX nasal airway dysfunction and a second active agent comprising an
XX antiinflammatory steroid and ubiqunone. A composition of the invention
XX has antiinflammatory, antiallergic, antiasthmatic, hypotensive,
XX immunosuppressive, and cyostatic activity. The composition may have a
XX use in antisense gene therapy. The composition is useful for treating or
XX preventing a respiratory, lung or malignant disease or condition, also
XX for enhancing the prophylactic or therapeutic respiratory effect of an
XX antiinflammatory steroid in a subject, for reducing or depleting levels
XX of, or reducing sensitivity to adenosine, reducing levels of adenosine
XX receptor, producing bronchodilation, increasing levels of ubiqunone or
XX lung surfactant in a subject's tissue, or treating bronchoconstriction,
XX lung inflammation, lung allergies, or a respiratory disease or condition.
XX Note: The sequence data for this patent is not represented in the printed
XX specification, but was obtained in electronic format directly from WIPO
XX at ftp.wipo.int/pub/published_pct_sequences

XX SQ Sequence 6771 BP; 1815 A; 1570 C; 1439 G; 1947 T; 0 U; 0 Other;
Query Match 85.7%; Score 429.2; DB 10; Length 6771;
Best Local Similarity 97.4%; Pred. No. 1.3e-112;
Matches 489; Conservative 0; Mismatches 8; Indels 5; Gaps 5;
Qy 1 ACGGGGCTCACTCTGTGCTCCCGAGGCTGGAGTGGCGGTGATCTCAGATCACTACAAC 60
Db 1558 ACGGGGCTCACTCTGTGCTG-CCAGGCTGGAGTGGCGGTGATCTCAGATCACTACAAC 1616
Qy 61 CTCATCTCTGGGTTCAAATAATCTCTTGGCTCAGCCTCTCAAGTAGCTGGGACTTGT 120
Db 1617 CTCATCTCTCT-GGTTCAAATAATCTCTTGGCTCAGCCTCTCAAGTAGCT-GGACTTGT 1674
Qy 121 AGCTCTCAAGTAGCTGGCACACACCATGCCAGCTAAATTTTCTGCTGTTTTTTTG 180
Db 1675 AGCTCTCAAGTAGCTGGCACACACCATGCCAGCTAAATTTTCTGCTGTTTTTTTG 1734
Qy 181 GTAGAGACAGGTTTTTCAACCATGTTGGCCAGGCTGGGTGACCTCCCTTTTAGATTTCTCCTC 240
Db 1735 GTAGAGACAGGTTTTTCAACCATGTTGGCCAGGCTGGGTGACCTCCCTTTTAGATTTCTCCTC 1794
Qy 241 ATCTGCTCTATTTCTTCCCTTTCT-AATGAGATATCCAGTTTCTTCTTACATTCACATTT 299
Db 1795 ATCTGCTCTATTTCTTCCCTTTCTAAATGAGATATCCAGTTTCTTCTTACATTCACATTT 1854
Qy 300 ATTATTTATTTCTTATTATTATTTGACAGACTCTTGTCTCGCCAGGCTGGAGTACAG 359
Db 1855 ATTATTTATTTCTTATTATTATTTGACAGACTCTTGTCTTGTGCCAAGGCTGGAGTACAG 1914
Qy 360 TGGTGGCATCTCGGCTCACTGCAAGCTCCACCTGCTGGGTTCACGCCATTCTCCCGCTC 419
Db 1915 TGGTGGCATCTCGGCTCACTGCAAGCTCCACCTGCTGGGTTCACGCCATTCTCCCGCTC 1974
Qy 420 AGCTCTCCCGAGTAGCTGGGACTAAAGGCGCTGCCACCGCCCGCTAAATTTTGTGA 479
Db 1975 AGCTCTCCCGAGTAGCTGGGACTAAA-GCGCTGCCACCGCCCGCTAAATTTTGTGA 2033
Qy 480 TTTTAAATAAGACGGGGTTTC 501
Db 2034 TTTTAAATAAGACGGGGTTTC 2055
RESULT 11
ABD20768
ID ABD20768 standard; DNA; 6771 BP.
XX AC ABD20768;
XX DT 29-JUL-2004 (first entry)
XX DE Human pulmonary and inflammatory target DNA #379.
XX KW Human; antisense; bronchoconstriction; allergy; hyposecretion; pain;
KW respiratory tract inflammation; adenosine sensitivity; lung; cancer;
KW surfactant depletion; antiallergic; antiinflammatory; antiasthmatic;
KW analgesic; hypotensive; immunosuppressive; cyostatic; cystic fibrosis;
KW beta-adrenergic agonist; respiratory disease; pulmonary vasoconstriction;
KW respiratory distress syndrome; allergic rhinitis; pulmonary hypertension;
KW emphysema; chronic obstructive pulmonary disease; cancer; bronchitis;
KW pulmonary transplantation rejection; ds.
XX OS Homo sapiens.
XX WO200285309-A2.
XX 31-OCT-2002.
XX 23-APR-2002; 2002WO-US013143.
XX 24-APR-2001; 2001US-0286036P.
XX

PA	(EPIC-) EPIGENESIS PHARM INC.	
XX	Nyce JW, Li Y, Sandrasagra A, Katz E, Pabalan J, Aguilar D;	
PI	Miller S, Tang L, Shahabuddin S;	
XX	WPI; 2003-093058/08.	
DR	Pharmaceutical composition for treating asthma, has antisense	
XX	oligonucleotide containing less percentage of adenosine, targeted to	
PT	nucleic acids associated with lung airway or lung dysfunction, and	
PT	bronchodilating agent.	
XX	Claim 15; SEQ ID NO 12161; 763pp; English.	
PS	This invention describes a novel composition (a) a first active agent,	
XX	comprising oligonucleotides, effective for alleviating	
CC	bronchoconstriction, respiratory tract inflammation, allergies and	
CC	reducing adenosine sensitivity, levels of adenosine (A) or (A) receptors,	
CC	surfactant depletion or hyposecretion, when administered to a mammal. The	
CC	oligonucleotides are derived from a gene encoding or regulating	
CC	expression of a target polypeptide associated with lung airway or lung	
CC	dysfunction or cancer and can be anti-sense to the corresponding mRNA.	
CC	The invention also describes a kit, that comprises: (a) a delivery	
CC	device, in separate containers, (b) the oligonucleotides, (c)	
CC	instructions for adding a carrier and for use of the kit. The composition	
CC	of the invention has anti-allergic, anti-inflammatory, antiasthmatic,	
CC	analgesic, hypotensive, immunosuppressive and cytostatic activity, is a	
CC	beta-adrenergic agonist. The composition is useful for preventing or	
CC	treating a respiratory, lung or malignant disease. The administered	
CC	composition comprises oligo and is administered to reduce the production	
CC	or availability, or to increase the degradation of the target mRNA or to	
CC	reduce the amount of target polypeptide present in the lungs. The	
CC	pulmonary obstruction, and/or bronchoconstriction and/or lung	
CC	inflammation, allergies and/or surfactant hypoproduction are associated	
CC	with a disease or condition such as pulmonary vasoconstriction,	
CC	inflammation, allergies, asthma, impeded respiration, respiratory	
CC	distress syndrome, pain, cystic fibrosis, allergic rhinitis, pulmonary	
CC	hypertension, emphysema, chronic obstructive pulmonary disease, pulmonary	
CC	transplantation rejection, pulmonary infections, bronchitis or cancer.	
CC	The reduced adenosine content of the anti-sense oligos corresponding to	
CC	thymidines present in the target RNA serves to prevent the breakdown of	
CC	the oligonucleotides into products that free adenosine into the system	
CC	e.g., lung, brain, heart, kidney, etc. tissue environment and thereby, to	
CC	prevent any unwanted effects due to it	
XX	Sequence 6771 BP; 1815 A; 1570 C; 1439 G; 1947 T; 0 U; 0 Other;	
SQ		
Query Match 85.7%; Score 429.2; DB 11; Length 6771;		
Best Local Similarity 97.4%; Pred. No. 1.3e-112;		
Matches 489; Conservative 0; Mismatches 8; Indels 5; Gaps 5;		
QY	1	ACGGGGCTACTCTGTGCCAGCGTGGAGTGGGTGGGTGATCTCAGATCACTACAAC 60
Db	1558	ACGGGGCTACTCTGTGCG-CCAGCGTGGAGTGGGTGGGTGATCTCAGATCACTACAAC 1616
QY	61	CTCCATCTCTCTGGGTTCAATAATTCCTTGGCTCAGCCTCTCAAGTAGCTGGGACTTGT 120
Db	1617	CTCCATCTCT-GGTTCAATAATTCCTTGGCTCAGCCTCTCAAGTAGCT-GGACTTGT 1674
QY	121	AGCTCTCAAGTAGCTGGCACACACCAGTCCAGTCAATTTTGTGTTGTTTTTTT 180
Db	1675	AGCTCTCAAGTAGCTGGCACACACCAGTCAATTTTGTGTTGTTTTTTT 1734
QY	181	GTAGAGACAGTTTTCACCAATGTTGGCCAGGTGGGTGACTCCCTTTAGATTCTCCTC 240
Db	1735	GTAGAGACAGTTTTCACCAATGTTGGCCAGGTGGGTGACTCCCTTTAGATTCTCCTC 1794
QY	241	ATCCTGCTCTATTCTTCCCTTTCT-AATGCAGTATCCAGTTTCCTTACTTATCACTTT 299
Db	1795	ATCCTGCTCTATTCTTCCCTTTCTAATGCAGTATCCAGTTTCCTTACTTATCACTTT 1854
QY	300	ATTATTATCTTATTATTATGAGACAGATCTTGTTCGCCAAGGCTGGAGTACAG 359

Db	1855	ATTATTATCTTATTATTATTATGAGACAGATCTTGTTCGCCAAGGCTGGAGTACAG 1914
QY	360	TGTTGGATCTCGGCTCAGTCAAGCTCCACCTGCTGGGTTCACGCATTTCTCCCGCTC 419
Db	1915	TGTTGGATCTCGGCTCAGTCAAGCTCCACCTGCTGGGTTCACGCATTTCTCCCGCTC 1974
QY	420	AGCTCCCAAGTAGCTGGGACTAAAGCGCGCTGCCACCGCCCTCAATTTTTTGTGA 479
Db	1975	AGCTCCCAAGTAGCTGGGACTAAA-GCGCTGCCACCGCCCTCAATTTTTTGTGA 2033
QY	480	TTTTTAATAAAGACGGGGTTTC 501
Db	2034	TTTTTAATAAAGACGGGGTTTC 2055

RESULT 12		
ABK13076/c		
ID	ABK13076	standard; DNA; 35973 BP.
XX	AC	ABK13076;
XX	DT	23-APR-2002 (first entry)
XX	DE	Human amyloid beta precursor protein binding protein 1 gene.
XX	KW	Human; amyloid beta precursor protein binding protein 1; APPBP1; gene;
KW	Alzheimer's disease; transgenic animal; platelet aggregation;	
KW	single nucleotide polymorphism; SNP; ds.	
XX	OS	Homo sapiens.
XX	Key	Location/Qualifiers
FT	variation	replace(3436,G)
FT		/*tag= a
FT		/standard_name= "Single nucleotide polymorphism"
FT	variation	replace(3474,C)
FT		/*tag= b
FT		/standard_name= "Single nucleotide polymorphism"
FT	CDS	4085..31975
FT		/*tag= c
FT		/product= "APPBP1"
FT		/note= "Amyloid beta precursor protein binding protein 1"
FT	variation	replace(11314,G)
FT		/*tag= d
FT		/standard_name= "Single nucleotide polymorphism"
FT	variation	replace(18123,G)
FT		/*tag= e
FT		/standard_name= "Single nucleotide polymorphism"
FT	variation	replace(18571,A)
FT		/*tag= f
FT		/standard_name= "Single nucleotide polymorphism"
FT	variation	replace(21229,A)
FT		/*tag= g
FT		/standard_name= "Single nucleotide polymorphism"
FT	variation	replace(21312,A)
FT		/*tag= h
FT		/standard_name= "Single nucleotide polymorphism"
FT	variation	replace(24300,A)
FT		/*tag= i
FT		/standard_name= "Single nucleotide polymorphism"
FT	variation	replace(26093,A)
FT		/*tag= j
FT		/standard_name= "Single nucleotide polymorphism"
FT	variation	replace(29334,C)
FT		/*tag= k
FT		/standard_name= "Single nucleotide polymorphism"
FT	variation	replace(31499,T)
FT		/*tag= l
FT		/standard_name= "Single nucleotide polymorphism"
FT	variation	replace(31582,G)
FT		/*tag= m
FT		/standard_name= "Single nucleotide polymorphism"
XX		

```
PN WO200202820-A1.
XX 10-JAN-2002.
XX 02-JUL-2001; 2001WO-US020951.
XX 30-JUN-2000; 2000US-0215511P.
XX (GENA-) GENAISSANCE PHARM INC.
XX Anastasio AE, Chew A, Choi JY, Kazemi A, Koshy B, Sausker EA;
PI Stephens CJ,
XX WPI; 2002-164539/21.
DR P-PSDB; AAU11030.
XX
XX Amyloid beta precursor protein binding protein 159 kD (APPBP1) gene
PT polymorphic variants, useful e.g. in studying the expression and function
PT of APPBP1 and screening candidate drugs for treating Alzheimer's disease.
XX
XX Claim 1; Fig 1; 104pp; English.
XX
XX The invention relates to an isolated polypeptide comprising a sequence
CC which is a polymorphic variant of a reference sequence for the amyloid
CC beta precursor protein binding protein 1, 59kD (APPBP1) protein or its
CC fragment. The polymorphic variants are useful in studying the expression
CC and function of APPBP1, in expressing APPBP1 protein for use in screening
CC for candidate drugs to treat diseases related to APPBP1 activity, in
CC studying the effect of the variation on the biological activity of
CC APPBP1, and the binding affinity of candidate drugs targeting APPBP1 for
CC the treatment of disorders such as Alzheimer's disease. The haplotyping
CC methods are useful in validating APPBP1 as a candidate target for
CC treating a specific condition or disease predicted to be associated with
CC APPBP1 activity, or in the design of clinical trials of candidate drugs
CC for treating a specific condition or disease associated with APPBP1
CC activity. The transgenic animals are useful for studying expression of
CC the APPBP1 isogenes in vivo, for in vivo screening and testing of drugs
CC targeted against APPBP1 protein, and for testing the efficacy of
CC therapeutic agents and compounds for disorders related to platelet
CC aggregation in a biological system. The present sequence represents the
CC human APPBP1 gene as described in the method of the invention
XX
XX Sequence 35973 BP; 9909 A; 7327 C; 7959 G; 10766 T; 0 U; 12 Other;
SQ
Query Match 46.2%; Score 231.6; DB 6; Length 35973;
Best Local Similarity 72.4%; Pred No. 1.6e-55;
Matches 349; Conservative 0; Mismatches 114; Indels 19; Gaps 3;
QY 20 CCAGCTGGAGTGGCGGTGATCTCAGATCACTACAACTCCATCTCTGGGTCAA 79
DB 22424 CCAGCTGGAGTGGCGGACGATCTTGGCTCACTGCAACTCGTCTCTGGGTCAA 22365
QY 80 ATAATCTTGGCTCAGCTCTCAAGTAGCTGGGACTTAGCTCTCAAGTAGTGCA 139
DB 22364 GCGATTCCCTCGCTCAGCTCCCGAATAGTGGACT-----ACAGGTG 22320
QY 140 CACACACATGCCCGCAGCTAATTTTGTGTGTTTTTGTGTAGACAGGTTTTCAC 199
DB 22319 CGCACACACACACCTGGCTAATTTTGT-----ATTTTGTAGTAGACAGGTTTTCAC 22263
QY 200 ATGTTGGCAGCTGGGTGACCTCCCTTTTGTAGATTTCTCTCATCTGCTATTCTTCCC 259
DB 22262 ATGTTGGCAGGAT--GGTCTTGATCTTGTGACCTTGATCCACCGTCTCGGCTCCG 22204
QY 260 CTTTCTAATGAGTATCAGTTTCTTACTTATACATATTTATTTATTTATTTAT 319
DB 22203 AAGTGGTGGGATTAACAGCGGTGAGCCACCGCCCGGCTAAGTTCTTTTTTTTTTT 22144
QY 320 TGAGACAGATCTTGTGTCGCAAGCTGGAGTACAGTGGTGGATCTCGGCTCACT 379
DB 22143 TGAGACAGATCTCACTCTGTGCGCCAGCTGGAGTGGAGTGGGATCTCGGCTCACT 22084
QY 380 GCAAGCTCCACTGTGTGGTTTCAGGCCATTCTCCCGCCTCAGCCTCCCGCAGTAGCTGGGA 439
DB 22083 GCAAGCTCCGCTCCCGGTTTCAAGCTATTCTCTGCTCAGCTCCCGGCTAGCTGGGA 22024
QY 440 CTAAAGGCGCTGCCACACGCGCGCTAATTTTGTATTTTAAATAAGACGGGTT 499
DB 22023 CTACAGGCGCTGCCACACGCGCGCTAATTTTGTATTTTAAATAAGACGGGTT 21964
QY 500 TC 501
DB 21963 TC 21962
XX
XX RESULT 13
XX AAL06681/c
XX ID AAL06681 standard; DNA; 602 BP.
XX AC AAL06681;
XX DT 21-NOV-2001 (first entry)
XX DE Human reproductive system related antigen DNA SEQ ID NO: 9369.
XX KW Human; reproductive system related antigen; reproductive system disorder;
XX KW cancer; gene therapy; ds.
XX OS Homo sapiens.
XX PN WO200155320-A2.
XX PD 02-AUG-2001.
XX PF 17-JAN-2001; 2001WO-US001339.
XX PR 31-JAN-2000; 2000US-0179065P.
XX PR 04-FEB-2000; 2000US-0180628P.
XX PR 24-FEB-2000; 2000US-0184664P.
XX PR 02-MAR-2000; 2000US-0186350P.
XX PR 16-MAR-2000; 2000US-0189874P.
XX PR 17-MAR-2000; 2000US-0190076P.
XX PR 18-APR-2000; 2000US-0198123P.
XX PR 19-MAY-2000; 2000US-0205515P.
XX PR 07-JUN-2000; 2000US-0209467P.
XX PR 28-JUN-2000; 2000US-0214886P.
XX PR 30-JUN-2000; 2000US-0215135P.
XX PR 07-JUL-2000; 2000US-0216647P.
XX PR 07-JUL-2000; 2000US-0216880P.
XX PR 11-JUL-2000; 2000US-0217487P.
XX PR 11-JUL-2000; 2000US-0217496P.
XX PR 14-JUL-2000; 2000US-0218290P.
XX PR 26-JUL-2000; 2000US-0220963P.
XX PR 26-JUL-2000; 2000US-0220964P.
XX PR 14-AUG-2000; 2000US-0224518P.
XX PR 14-AUG-2000; 2000US-0224519P.
XX PR 14-AUG-2000; 2000US-0225213P.
XX PR 14-AUG-2000; 2000US-0225214P.
XX PR 14-AUG-2000; 2000US-0225266P.
XX PR 14-AUG-2000; 2000US-0225267P.
XX PR 14-AUG-2000; 2000US-0225268P.
XX PR 14-AUG-2000; 2000US-0225270P.
XX PR 14-AUG-2000; 2000US-0225447P.
XX PR 14-AUG-2000; 2000US-0225757P.
XX PR 14-AUG-2000; 2000US-0225758P.
XX PR 14-AUG-2000; 2000US-0225759P.
XX PR 18-AUG-2000; 2000US-0226279P.
XX PR 22-AUG-2000; 2000US-0226681P.
XX PR 22-AUG-2000; 2000US-0226686P.
XX PR 23-AUG-2000; 2000US-0227182P.
XX PR 23-AUG-2000; 2000US-0227009P.
XX PR 30-AUG-2000; 2000US-0228924P.
XX PR 01-SEP-2000; 2000US-0229287P.
XX PR 01-SEP-2000; 2000US-0229343P.
XX PR 01-SEP-2000; 2000US-0229344P.
XX PR 01-SEP-2000; 2000US-0229345P.
```

PR	05-SEP-2000;	2000US-0229509P.	PR	17-NOV-2000;	2000US-0249216P.
PR	05-SEP-2000;	2000US-0229513P.	PR	17-NOV-2000;	2000US-0249217P.
PR	06-SEP-2000;	2000US-0230437P.	PR	17-NOV-2000;	2000US-0249218P.
PR	06-SEP-2000;	2000US-0230438P.	PR	17-NOV-2000;	2000US-0249244P.
PR	08-SEP-2000;	2000US-0231242P.	PR	17-NOV-2000;	2000US-0249245P.
PR	08-SEP-2000;	2000US-0231243P.	PR	17-NOV-2000;	2000US-0249265P.
PR	08-SEP-2000;	2000US-0231244P.	PR	17-NOV-2000;	2000US-0249297P.
PR	08-SEP-2000;	2000US-0231413P.	PR	17-NOV-2000;	2000US-0249299P.
PR	08-SEP-2000;	2000US-0231414P.	PR	17-NOV-2000;	2000US-0249300P.
PR	08-SEP-2000;	2000US-0232080P.	PR	01-DEC-2000;	2000US-0250160P.
PR	08-SEP-2000;	2000US-0232082P.	PR	01-DEC-2000;	2000US-0250391P.
PR	12-SEP-2000;	2000US-0231968P.	PR	05-DEC-2000;	2000US-0251030P.
PR	14-SEP-2000;	2000US-0232397P.	PR	05-DEC-2000;	2000US-0251988P.
PR	14-SEP-2000;	2000US-0232398P.	PR	06-DEC-2000;	2000US-0256719P.
PR	14-SEP-2000;	2000US-0232399P.	PR	08-DEC-2000;	2000US-0251856P.
PR	14-SEP-2000;	2000US-0232400P.	PR	08-DEC-2000;	2000US-0251868P.
PR	14-SEP-2000;	2000US-0232401P.	PR	08-DEC-2000;	2000US-0251869P.
PR	14-SEP-2000;	2000US-0233063P.	PR	08-DEC-2000;	2000US-0251989P.
PR	14-SEP-2000;	2000US-0233064P.	PR	08-DEC-2000;	2000US-0251990P.
PR	21-SEP-2000;	2000US-0233065P.	PR	11-DEC-2000;	2000US-0254097P.
PR	21-SEP-2000;	2000US-0234223P.	PR	05-JAN-2001;	2001US-0255678P.
PR	21-SEP-2000;	2000US-0234274P.	XX		
PR	25-SEP-2000;	2000US-0234997P.	XX	(HUMA-) HUMAN GENOME SCI INC.	
PR	25-SEP-2000;	2000US-0234998P.	XX		
PR	26-SEP-2000;	2000US-0234999P.	XX	Rosen CA, Barash SC, Ruben SM;	
PR	27-SEP-2000;	2000US-0235484P.	XX	WPI; 2001-465570/50.	
PR	27-SEP-2000;	2000US-0235834P.	XX		
PR	29-SEP-2000;	2000US-0235836P.	XX		
PR	29-SEP-2000;	2000US-0236327P.	XX		
PR	29-SEP-2000;	2000US-0236367P.	XX		
PR	29-SEP-2000;	2000US-0236368P.	XX		
PR	29-SEP-2000;	2000US-0236369P.	XX		
PR	29-SEP-2000;	2000US-0236370P.	XX		
PR	02-OCT-2000;	2000US-0236802P.	XX		
PR	02-OCT-2000;	2000US-0237037P.	XX		
PR	02-OCT-2000;	2000US-0237038P.	XX		
PR	02-OCT-2000;	2000US-0237039P.	XX		
PR	02-OCT-2000;	2000US-0237040P.	XX		
PR	13-OCT-2000;	2000US-0239935P.	XX		
PR	13-OCT-2000;	2000US-0239937P.	XX		
PR	20-OCT-2000;	2000US-0240960P.	XX		
PR	20-OCT-2000;	2000US-0241221P.	XX		
PR	20-OCT-2000;	2000US-0241785P.	XX		
PR	20-OCT-2000;	2000US-0241786P.	XX		
PR	20-OCT-2000;	2000US-0241787P.	XX		
PR	20-OCT-2000;	2000US-0241808P.	XX		
PR	20-OCT-2000;	2000US-0241809P.	XX		
PR	20-OCT-2000;	2000US-0241826P.	XX		
PR	01-NOV-2000;	2000US-0244617P.	XX		
PR	08-NOV-2000;	2000US-0246474P.	XX		
PR	08-NOV-2000;	2000US-0246475P.	XX		
PR	08-NOV-2000;	2000US-0246476P.	XX		
PR	08-NOV-2000;	2000US-0246477P.	XX		
PR	08-NOV-2000;	2000US-0246478P.	XX		
PR	08-NOV-2000;	2000US-0246523P.	XX		
PR	08-NOV-2000;	2000US-0246524P.	XX		
PR	08-NOV-2000;	2000US-0246525P.	XX		
PR	08-NOV-2000;	2000US-0246526P.	XX		
PR	08-NOV-2000;	2000US-0246527P.	XX		
PR	08-NOV-2000;	2000US-0246528P.	XX		
PR	08-NOV-2000;	2000US-0246532P.	XX		
PR	08-NOV-2000;	2000US-0246609P.	XX		
PR	08-NOV-2000;	2000US-0246610P.	XX		
PR	08-NOV-2000;	2000US-0246611P.	XX		
PR	17-NOV-2000;	2000US-0249207P.	XX		
PR	17-NOV-2000;	2000US-0249208P.	XX		
PR	17-NOV-2000;	2000US-0249209P.	XX		
PR	17-NOV-2000;	2000US-0249210P.	XX		
PR	17-NOV-2000;	2000US-0249211P.	XX		
PR	17-NOV-2000;	2000US-0249212P.	XX		
PR	17-NOV-2000;	2000US-0249213P.	XX		
PR	17-NOV-2000;	2000US-0249214P.	XX		
PR	17-NOV-2000;	2000US-0249215P.	XX		

Isolated nucleic acid molecule encoding a reproductive system antigen is used in preventing, treating or ameliorating a medical condition.

Disclosure; SEQ ID NO 9369; 1297bp + Sequence Listing; English.

The present invention provides the protein and coding sequences of a number of human reproductive system related antigens. These can be used in the prevention and treatment of reproductive system disorders, including cancer. The present sequence is a genomic sequence encoding a protein of the invention

Query Match	46.2%;	Score 231.4;	DB 4;	Length 602;
Best Local Similarity	72.3%;	Pred. No. 3.7e-56;		
Matches 362;	Conservative 0;	Mismatches 121;	Indels 18;	Gaps 4;
QY	3	GGGGCTCACTCTGTCCGCGAGCTGGAGTGGCGTGGCGTGGATCTCAGATCACTACAACCT	62	
Db	571	GAGTCTTGTCTGCTAGCTAGCTGGAGTGGCGGATCTCGGGTCACTGCAACCT	512	
QY	63	CCATCTCTGGGTTCAATAATTCCTTGGCTCAGCCTCTCAAGTAGCTGGGACTGTAG	122	
Db	511	CGGCTCCGGGTTACGGCAATCTCTGCTCAGCCTCCGAGTAGTGGGACT-----	457	
QY	123	CTCTCAGTAGTGGCACACACACATGCCAGCTAAATTTTGT- GTTTTTTTGG	181	
Db	456	-----ACAGGCGCCACACACCGCCGCTAAATTTTGTATTTTTTTTAG	407	
QY	182	TAGAGACAGTTTACCATGTTGGCCAGCTGGGTGACCTCCCTTTTAGATTCTCCTCA	241	
Db	406	TAGAGACAGTTTACCGTGTAGCCAGATGGTCTCGATCTCCGACCTCGTGGTCCG	347	
QY	242	TCCTGCTC-TATTTCTCCCTTTCTAATGCAATATCCAGTTCTTCTTACTATACATTTA	300	
Db	346	CCGCGCTCGGCTCCGAAAGTGTGGGATTACAGGCGTGGGACCTCGCCCGGCTTTT	287	
QY	301	TTATTTATTTTATTATTATGACAGATCTTTGCTTGTCCGCAAGGCTGGAGTACAGT	360	
Db	286	TTTTTTTTTTTTTTTTTTTGTAGACAGATCTTGTCTGTCCCGAGGCTGGAGTGCAGT	227	
QY	361	GGTGGATCTCGGCTCACTGCAAGCTCCACCTGCTGGGTTACGCCATTCTCCGCTCA	420	

Db 226 GGGCGGATCTGGCTCACTGCAAGCTCCACCTCTGGGTTTCAACCAATCTCTCGCTCA 167
Qy 421 GCTCCCGAGTACTGGGACTAAAGCGGCTGCGACCGCCGCTAAATTTTGTAT 480
Db 166 GCTCCCGAGTACTGGGACTACAGCGGCTGCCACCATGCTGCTAA-TTTTGTAT 108
Qy 481 TTTTAATAAGACGGGGTTTC 501
Db 107 TTTTAGTAGACGGGGTTTC 87

RESULT 14
ABA08020/c
ID ABA08020 standard; DNA; 602 BP.
XX
AC ABA08020;
DT
XX 11-JAN-2002 (first entry)
DE Human ovarian and breast cancer associated polynucleotide SEQ ID NO 815.
KW Cytostatic; immunosuppressive; nootropic; neuroprotective; antiviral;
KW anti-allergic; hepatotropic; antidiabetic; anti-inflammatory; antitumor;
KW vulvar; anticonvulsant; antibacterial; antifungal; antiparasitic;
KW cardiac; gene therapy; cancer; immune disorder; cardiovascular disorder;
KW neurological disease; infection; human; secreted protein; ds.
OS Homo sapiens.
XX
XX WO20015325-A2.
PD
PF 02-AUG-2001.
PF
XX 17-JAN-2001; 2001WO-US001345.
XX
PR 31-JAN-2000; 2000US-0179065P.
PR 04-FEB-2000; 2000US-0180628P.
PR 24-FEB-2000; 2000US-0184664P.
PR 02-MAR-2000; 2000US-0186350P.
PR 16-MAR-2000; 2000US-0189874P.
PR 17-MAR-2000; 2000US-0190078P.
PR 18-APR-2000; 2000US-0198123P.
PR 19-MAY-2000; 2000US-0205515P.
PR 07-JUN-2000; 2000US-0209467P.
PR 28-JUN-2000; 2000US-0214886P.
PR 30-JUN-2000; 2000US-0215135P.
PR 07-JUL-2000; 2000US-0216647P.
PR 07-JUL-2000; 2000US-0216880P.
PR 11-JUL-2000; 2000US-0217487P.
PR 11-JUL-2000; 2000US-0217496P.
PR 26-JUL-2000; 2000US-0220963P.
PR 26-JUL-2000; 2000US-0220964P.
PR 14-AUG-2000; 2000US-0224518P.
PR 14-AUG-2000; 2000US-0224519P.
PR 14-AUG-2000; 2000US-0225213P.
PR 14-AUG-2000; 2000US-0225214P.
PR 14-AUG-2000; 2000US-0225266P.
PR 14-AUG-2000; 2000US-0225267P.
PR 14-AUG-2000; 2000US-0225268P.
PR 14-AUG-2000; 2000US-0225270P.
PR 14-AUG-2000; 2000US-0225447P.
PR 14-AUG-2000; 2000US-0225757P.
PR 14-AUG-2000; 2000US-0225758P.
PR 14-AUG-2000; 2000US-0225759P.
PR 18-AUG-2000; 2000US-0226279P.
PR 22-AUG-2000; 2000US-0226681P.
PR 22-AUG-2000; 2000US-0226686P.
PR 23-AUG-2000; 2000US-0227182P.
PR 23-AUG-2000; 2000US-0227009P.
PR 30-AUG-2000; 2000US-0228924P.
PR 01-SEP-2000; 2000US-0229287P.
PR 01-SEP-2000; 2000US-0229343P.
PR 01-SEP-2000; 2000US-0229344P.
PR 01-SEP-2000; 2000US-0229345P.
PR 05-SEP-2000; 2000US-0229509P.
PR 05-SEP-2000; 2000US-0229513P.
PR 06-SEP-2000; 2000US-0230437P.
PR 06-SEP-2000; 2000US-0230438P.
PR 08-SEP-2000; 2000US-0231242P.
PR 08-SEP-2000; 2000US-0231243P.
PR 08-SEP-2000; 2000US-0231244P.
PR 08-SEP-2000; 2000US-0231413P.
PR 08-SEP-2000; 2000US-0231414P.
PR 08-SEP-2000; 2000US-0232080P.
PR 08-SEP-2000; 2000US-0232081P.
PR 12-SEP-2000; 2000US-0231968P.
PR 14-SEP-2000; 2000US-0232397P.
PR 14-SEP-2000; 2000US-0232398P.
PR 14-SEP-2000; 2000US-0232399P.
PR 14-SEP-2000; 2000US-0232400P.
PR 14-SEP-2000; 2000US-0232401P.
PR 14-SEP-2000; 2000US-0233063P.
PR 14-SEP-2000; 2000US-0233064P.
PR 14-SEP-2000; 2000US-0233065P.
PR 21-SEP-2000; 2000US-0234223P.
PR 21-SEP-2000; 2000US-0234274P.
PR 25-SEP-2000; 2000US-0234997P.
PR 25-SEP-2000; 2000US-0234998P.
PR 26-SEP-2000; 2000US-0235484P.
PR 27-SEP-2000; 2000US-0235834P.
PR 27-SEP-2000; 2000US-0235836P.
PR 29-SEP-2000; 2000US-0236327P.
PR 29-SEP-2000; 2000US-0236367P.
PR 29-SEP-2000; 2000US-0236368P.
PR 29-SEP-2000; 2000US-0236369P.
PR 29-SEP-2000; 2000US-0236370P.
PR 02-OCT-2000; 2000US-0236802P.
PR 02-OCT-2000; 2000US-0237037P.
PR 02-OCT-2000; 2000US-0237038P.
PR 02-OCT-2000; 2000US-0237039P.
PR 13-OCT-2000; 2000US-0237040P.
PR 13-OCT-2000; 2000US-0239935P.
PR 13-OCT-2000; 2000US-0239937P.
PR 20-OCT-2000; 2000US-0240960P.
PR 20-OCT-2000; 2000US-0241221P.
PR 20-OCT-2000; 2000US-0241785P.
PR 20-OCT-2000; 2000US-0241786P.
PR 20-OCT-2000; 2000US-0241787P.
PR 20-OCT-2000; 2000US-0241808P.
PR 20-OCT-2000; 2000US-0241809P.
PR 20-OCT-2000; 2000US-0241826P.
PR 01-NOV-2000; 2000US-0244617P.
PR 08-NOV-2000; 2000US-0246474P.
PR 08-NOV-2000; 2000US-0246475P.
PR 08-NOV-2000; 2000US-0246476P.
PR 08-NOV-2000; 2000US-0246477P.
PR 08-NOV-2000; 2000US-0246478P.
PR 08-NOV-2000; 2000US-0246523P.
PR 08-NOV-2000; 2000US-0246524P.
PR 08-NOV-2000; 2000US-0246525P.
PR 08-NOV-2000; 2000US-0246526P.
PR 08-NOV-2000; 2000US-0246527P.
PR 08-NOV-2000; 2000US-0246528P.
PR 08-NOV-2000; 2000US-0246532P.
PR 08-NOV-2000; 2000US-0246609P.
PR 08-NOV-2000; 2000US-0246610P.
PR 08-NOV-2000; 2000US-0246611P.
PR 17-NOV-2000; 2000US-0246613P.
PR 17-NOV-2000; 2000US-0249207P.
PR 17-NOV-2000; 2000US-0249208P.
PR 17-NOV-2000; 2000US-0249209P.
PR 17-NOV-2000; 2000US-0249210P.
PR 17-NOV-2000; 2000US-0249211P.
PR 17-NOV-2000; 2000US-0249212P.
PR 17-NOV-2000; 2000US-0249213P.

CC The sequence data for this patent did not form part of the printed
 CC specification, but was obtained in electronic format directly from WIPO
 CC at ftp.wipo.int/pub/published_pat_sequences
 XX
 SQ Sequence 127508 BP; 33168 A; 27130 C; 29586 G; 37624 T; 0 U; 0 Other;

Query Match 45.7%; Score 228.8; DB 13; Length 127508;
 Best Local Similarity 70.8%; Pred. No. 1.7e-54;
 Matches 358; Conservative 0; Mismatches 122; Indels 26; Gaps 3;

QY 3 GGGGCTCACTGTGCGCCAGGCTGAGTGGCGGTGATCTCAGATCACTACAACCT 62
 DB |||||
 QY 106050 GAGTCTCGCTCTGTACCCAGGCTGAGTGCAGTGTGGGATCTCGGCTCACTGCAAGCT 106109
 DB |||||

QY 63 CCATCTCCCTGGGTTCAAAATAATTCTCTTGGCTCAGCTCTCAAGTAGCTGGGACTTGTAG 122
 DB |||||
 QY 106110 CGGCTCCCGGTTTCAACCCATTCTCTGCTCAGCTCCCGAGTAGCTGGGACT----- 106164
 DB |||||

QY 123 CTCTCAAGTAGCTGGGACACACACACATGCCAGCTAATTTTGTGTGTGTGTGTGTGTGT 182
 DB |||||

QY 106165 -----ACAGGCGGCCACACCATGCCAGCTAATTTT-----TTGTATTTTAGT 106210
 DB |||||

QY 183 AGAGACAGGTTTTCACCATGTTGGCCAGGCTGGGTGACCTCCCTT-----TTAGATTC 235
 DB |||||

QY 106211 AGAGACAGGTTTTCACCTGTGTAGCCAGGATGATCTCGATCTCCTGACCTGGGATCCGC 106270
 DB |||||

QY 236 TCCTCATCTGCTCTATTCTTCCCTTTCTAATGAGTATCCAGTTTCCTTACTTATCAC 295
 DB |||||

QY 106271 CGTCTTGGCTCCCAAGTGTGGGATTACAGGCGTGAGCCACACGTCGCGCCTTTT 106330
 DB |||||

QY 296 ATTTATTATTATTCTTATTATTATTGAGACAGAGTCTTGTCTTTGTGCGCAAGGCTGGAGT 355
 DB |||||

QY 106331 TTGCTTTTGTGTTTGTGTTTGTGTTTGTGTTTGTGTTTGTGTTTGTGTTTGTGTTTGT 106390
 DB |||||

QY 356 ACAGTGTGGGCTCTCGGCTCACTGCAAGCTCCACTGCTGGGTTTACGCCATTCTCCG 415
 DB |||||

QY 106391 GCAGTGGGCGCAATCTTGGCTCACTGCAAGCTCCGCTTCCGGGTTTCCGGGCAATCTCCTG 106450
 DB |||||

QY 416 CCTCAGCTCCCGAGTAGCTGGGACTAAAGGCGCTGCGCACCCAGCCCGCTAATTTT 475
 DB |||||

QY 106451 CCTCAGCTCCCGAGTAGCTGGGACTACAGGCGCCGCGCACCGCGCCCGGCTAATTTT 106510
 DB |||||

QY 476 TGTATTTTAAAGACGGGTTTC 501
 DB |||||

QY 106511 TGTATTTTGTAGAGACGGGTTTC 106536
 DB |||||

Search completed: February 9, 2006, 13:50:53
 Job time : 346.959 secs

THIS PAGE BLANK (USPTO)

GenCore version 5.1.1.7
Copyright (c) 1993 - 2006 Bioacceleration Ltd.

OM nucleic - nucleic search, using sw model

Run on: February 9, 2006, 13:34:47 ; Search time 2363.65 Seconds
(without alignments)
9917.025 Million cell updates/sec

Title: US-10-607-806-1_COPY_5000_5500

Perfect score: 501

Sequence: 1 acg9gggtcaactctgtcgcc.....tttaataaagacg9gggttc 501

Scoring table: IDENTITY_NUC

Gapop 10.0 , Gapext 1.0

Searched: 41078325 seqs, 23393541228 residues

Total number of hits satisfying chosen parameters: 82156650

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database :

EST:*

1: gb_est1:*

2: gb_est2:*

3: gb_est3:*

4: gb_hic:*

5: gb_est4:*

6: gb_est5:*

7: gb_est6:*

8: gb_est7:*

9: gb_gsa1:*

10: gb_gsa2:*

11: gb_gsa3:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	213.2	42.6	810	9	BZ609826 WHABQ60TR
C 2	213.2	42.6	877	6	CD515864
C 3	211.6	42.2	627	6	CD702127
4	211.4	42.2	759	1	AW976010
C 5	210.6	42.0	820	5	BU568843
6	210.6	42.0	3775	4	CR627381
7	209.8	41.9	689	9	BZ605978
C 8	209.6	41.8	642	5	BX496786
9	209.4	41.8	631	7	CK002773
C 10	208.4	41.6	623	6	CD723982
11	208.2	41.6	891	5	BU174155
12	208.2	41.6	1654	4	CR591254
13	207.4	41.4	777	8	DR761502
14	207.4	41.4	832	5	BU180590
15	207.4	41.4	935	5	BU838401
16	206.6	41.2	620	3	BP382151
17	206.6	41.2	771	5	BQ776208
18	206	41.1	671	3	BM722145
C 19	206	41.1	782	5	BX644719
20	205.6	41.0	520	5	BQ807308
21	204.4	40.8	618	3	BP382202
C 22	204.2	40.8	613	9	AQ419352

23	204.2	40.8	630	1	AI457389
C 24	203.6	40.6	787	8	CK786428
C 25	203	40.5	2128	4	CR925964
C 26	201.6	40.2	908	5	BU171410
27	201.4	40.2	685	2	BU193100
28	201.4	40.2	1014	5	BQ709843
C 29	200.6	40.0	743	8	DR759591
C 30	200.2	40.0	714	8	CK872940
C 31	199.2	39.8	990	1	AW600804
C 32	198.6	39.6	718	7	CN359768
C 33	198.4	39.5	732	2	BG542372
34	198	39.5	512	1	AL135377
C 35	197.6	39.4	538	9	AQ585309
C 36	197.6	39.4	877	9	AQ739838
C 37	197.4	39.4	942	9	AQ743787
C 38	197	39.3	621	9	AQ636660
C 39	196.2	39.2	687	10	AG165009
C 40	195.8	39.1	910	5	BQ722917
C 41	195.6	39.0	586	1	AL701696
42	195.4	39.0	576	5	BX508289
C 43	195.4	39.0	666	9	AQ342859
C 44	195.4	39.0	1921	4	AF370377
C 45	195.2	39.0	493	2	BF826830

ALIGNMENTS

RESULT 1
BZ609826
LOCUS
DEFINITION
BZ609826 WHABQ60TR Human MCF7 breast cancer cell line library (MCF7.1) Homo sapiens genomic clone MCF7_1-11123, genomic survey sequence.
ACCESSION
BZ609826
VERSION
BZ609826.1 GI:31518387
KEYWORDS
Homo sapiens (human)
ORGANISM
Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini; Hominidae; Homo.
REFERENCE
1 (bases 1 to 810)
Vollik,S., Zhao,S., Chin,K., Brebner,J.H., Herndon,D.R., Tao,Q., Kowbel,D., Huang,G., Lapuk,A., Kuo,W.-L., Magrane,G., de Jong,P., Gray,J.W. and Collins,C.
End-sequence profiling: Sequence-based analysis of aberrant genomes
Proc. Natl. Acad. Sci. U.S.A. 100 (13), 7696-7701 (2003)
CONTACT: Volik SV
Colin Collins' lab
UCSF Comprehensive Cancer Center
UCSF Box 0808, San Francisco, CA 94143-0808, USA
Tel: 415 502 7066
Fax: 415 502 5665
Email: svolik@cc.ucsf.edu
This clone is available from Amplicon Express
http://www.genomex.com
Class: BAC ends.
Location/Qualifiers
1. .810
/organism="Homo sapiens"
/mol_type="genomic DNA"
/db_xref="taxon:9606"
/clone="MCF7_1-11123"
/sex="female"
/clone_lib="Human MCF7 breast cancer cell line library (MCF7.1)"
/notes="Vector: pECBAC1; Site 1: HindIII; This library was constructed from MCF7 breast cancer cell line by Amplicon Express (http://www.genomex.com) using their standard procedure."

ORIGIN

JOURNAL
COMMENT

Unpublished (2003)
Contact: Yixin Zeng
Cancer Center
Sun Yat-sen University
651 Dongfeng Road East, Guangzhou 510060, China
Tel: 86-1380-9770-743
Fax: 86-20-8775-4506
Email: yxzeng@zsus.edu.cn.

FEATURES
source

1. .627
Location/Qualifiers
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/tissue_type="normal nasopharynx"
/clone_lib="human nasopharynx"
/note="ESTs generated from a normal nasopharynx cdna library from southern Chinese"

ORIGIN

Query Match 42.2%; Score 211.6; DB 6; Length 627;
Best Local Similarity 69.8%; Pred. No. 1.4e-30;
Matches 359; Conservative 0; Mismatches 119; Indels 36; Gaps 4;

```

QY 3 GGGGCTCACTCTGTGCGCCAGGCTGGAGTGGCGGTGATCTCAGATCACTCAACCT 62
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 623 GAGTCTCGCTCTGTGAGCCAGGCTGGAGTGCATGCGCGATCTCAGCTCACTGCAACCT 564

QY 63 CCATCTCTCTGGGTTCAAATAATTTCTTTCCTCAGCTCTCAAGTAGTGGAGCTTGAG 122
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 563 CTGCTCTCCAGGTTCAAGCCATCTCTGCTCAGCCACCTGAGTAGCTGGGA----- 511

QY 123 CTCTCAAGTAGTGGCAGACACACCATGCGCCAGCTAAATTTTGTGTGTTTTTTTGGT 182
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 510 -----TTACAGGTGGCTCAAAGACACCCAGCTAAATTTT-----GTATTTTGTAGT 465

QY 183 AGAGACAGGTTTTCACCATGTTGGCCAGGCTGG-----GTGACCTCCCTTTTAGA 232
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 464 AGAGACAGGTTTTCACCTGTGTGTTGGCCAGGCTGGTCTCAAACTCTAAGCTCAGGTTATCC 405

QY 233 TTCTCTCATCTGCTCTATTCTTCCCTTTCTTAATGAGTATCCAGTTTCTCTACTTA- 291
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 404 ACCTGCTCGGCTCACAGAGTGTGGGANTACAGAGTAGCCACCGTGCCTGACCTAT 345

QY 292 ----TCACATTTATTATTCTTTATTATTATTTAGACAGAGTCTTGTCTTGCCCAAG 347
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 344 AGTTTTCAGTTTTTTTTTTTGTGTTTTTTTGTGAGCGAGTCTCGCTCTGTGCCAG 285

QY 348 GCTGAGTACAGTGTGGATCTCGGCTACCTGCAAGCTTCCACCTGCTGGGTTCAAGCCA 407
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 284 GCTGAGTACAGTGTGGATCTTGGCTCACTGCAAGCTCCACCTCTCTGGGTTCAAGCCA 225

QY 408 TTCTCCCGCTCAGCTCCCGAGTACTGGGACTAAAGCGCTGCGACCAAGCCCGCT 467
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 224 TTCTCTGCTCAGCTCCCAAGTAGTGGGACTACAGGCGCGCCGCCACCAAGCCCGCT 165

QY 468 AATTTTTTGTATTTTAAATAAGACGGGTTTC 501
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 164 AATTTTTTGTATTTTGTAGACACAGGGTTTC 131

```

RESULT 4

AW976010 759 bp mRNA linear EST 02-JUN-2000
LOCUS
DEFINITION EST388119 MAGE resequences, MAGN Homo sapiens cdna, mRNA sequence.
ACCESSION AW976010
VERSION AW976010.1 GI:8167232

KEYWORDS

EST.

SOURCE

Homo sapiens (human)

ORGANISM

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini;
Hominidae; Homo.
REFERENCE 1 (bases 1 to 759)

AUTHORS

Hegde, P., Qi, R., Abernathy, K., Dharap, S., Gaspard, R., Gay, C.,
Holt, I. E., Saeed, A. I., Sharov, V., Lee, N. H., Yeatman, T. J. and
Quackenbush, J.

TITLE

Assessment of gene expression patterns in a model of colon tumor
metastasis using a 19,200 element cDNA microarray

JOURNAL
COMMENT

Unpublished (2000)
Contact: John Quackenbush
The Institute for Genomic Research
9712 Medical Center Dr., Rockville, MD 20850, USA
Tel: 301 838 3528
Fax: 301 838 0208
Email: johnq@tigr.org
Plate: 356

FEATURES

source
Location/Qualifiers
1. .759
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/clone_lib="MAGE resequences, MAGN"
/note="Vector: pBluescriptSKm"

ORIGIN

```

Query Match 42.2%; Score 211.4; DB 1; Length 759;  
Best Local Similarity 70.0%; Pred. No. 1.4e-30;  
Matches 355; Conservative 0; Mismatches 121; Indels 31; Gaps 4;

QY 3 GGGGCTCACTCTGTGCGCCAGGCTGGAGTGGCGGTGATCTCAGATCACTCAACCT 62
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 131 GAGTCTCACTCTGTGACCCAGGCTGGAGTGCAGCGCATCTCGGCTCACTGCAAGCT 190

QY 63 CCATCTCTCTGGGTTCAAATAATTTCTTTCCTCAGCTCTCAAGTAGTGGAGCTTGAG 122
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 191 CCACCTCCCGGTTTCAGCCATCTCTGCTCAGCTTCCCGAGTAGCTGGGA----- 243

QY 123 CTCTCAAGTAGTGGCAGACACCATGCGCCAGCTAAATTTTGTGTGTTTTTTTGGT 182
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 244 -----TTACAGGCACATGCCACACGCGCGGCTACTTTT-----GTATTTTGTAGC 289

QY 183 AGAGACAGGTTTTCACCATGTTGGCCAGGCTGG-----GTGACCTCCCTTTTATAGT 233
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 290 AGAGACGGGTTTCACCATCTTGCCAGGCTGGTCTTGAGCTCTGACCTCGTATCCAC 349

QY 234 TCTCTCATCTGCTCTATTCTTCCCTTTCTTAATGAGTATCCAGTTTCTCTACTTATC 293
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 350 CCACCTCGGCTCCAAAGTGTGGGATTACAGCGGTAGCCACTGTGCTGGCTTGAAT 409

QY 294 ACATTATTATTATTCTTTATTATTATTATGAGACAGAGTCTTGTCTTGTGCGCAAGGCTGGA 353
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 410 TTTTCTTTTCTTTTCTTTTCTTTTGTGGAGCGAGTCTCGCTCTGCGCCAGGCTGGA 469

QY 354 GTACAGTGTGCGATCTCGGCTCACTGCAAGCTCCACCTGCTGGGTTTCAGCCATTCTCC 413
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 470 GTGCAATGGGCGCATCTCTCGGCTCACTGCAAGCTCCACCTCCCGGCTCCAGACCATCTCT 529

QY 414 CGCTCAGCTCCCGAGTCTGGAGCTAAAGCGCTGCCACACCGCCCGGCTAAATTTT 473
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 530 TGCCTCAGCTTCCCGAGTGTGGAGCTACAGAGCGCCGCCACCGCCCGGCTAA-TTT 588

QY 474 TTTGTATTTTAAATAAGACGGGTTT 500
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 589 TTTGTATTTTGTAGCAGACGGGTTT 615

```

RESULT 5

BU568843/c
LOCUS
DEFINITION AGENCOURT 10400218 NIH_MGC_82 Homo sapiens cdna clone IMAGE:6616116
5', mRNA sequence.

ACCESSION

BU568843

VERSION

BU568843.1

KEYWORDS

EST.

SOURCE

Homo sapiens (human)

ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini;
Hominidae; Homo.
1 (bases 1 to 820)
NIH-MGC http://mgc.nci.nih.gov/
National Institutes of Health, Mammalian Gene Collection (MGC)
Unpublished (1999)
Contact: Robert Strausberg, Ph.D.
Email: cgabbs-remail.nih.gov
Tissue Procurement: CLONTECH
CDNA Library Preparation: CLONTECH Laboratories, Inc.
CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
DNA Sequencing by: Agencourt Bioscience Corporation
Clone Distribution: MGC clone distribution information can be
found through the I.M.A.G.E. Consortium/LLNL at:
http://image.llnl.gov
Plate: LLM2859 row: e column: 12
High quality sequence stop: 556.
Location/Qualifiers
1. .820
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/clone="IMAGE:661616"
/lab_host="DH10B (T1 phage-resistant)"
/clone_lib="NIH MGC 82"
/note="Organ: testis; Vector: pDNR-LIB (Clontech); Site_1:
SfiI (ggccgctggcc); Site_2: SfiI (ggccattatggc); 5' and
3' adaptors were used in cloning as follows: 5' adaptor
sequence: 5'-CAGCGCATATGGCC-3' and 3' adaptor sequence:
5'-ATTCTAGAGCGGCGCGGCGCATG-dt(30)BN-3' (where B = A,
C, or G and N = A, C, G, or T). Average insert size
1.35 kb (range 0.9-4.0 kb). 14/15 colonies contained
inserts by PCR. This library was enriched for full-length
clones and was constructed by Clontech Laboratories (Palo
Alto, CA)."

FEATURES
source
1. .820
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/clone="IMAGE:661616"
/lab_host="DH10B (T1 phage-resistant)"
/clone_lib="NIH MGC 82"
/note="Organ: testis; Vector: pDNR-LIB (Clontech); Site_1:
SfiI (ggccgctggcc); Site_2: SfiI (ggccattatggc); 5' and
3' adaptors were used in cloning as follows: 5' adaptor
sequence: 5'-CAGCGCATATGGCC-3' and 3' adaptor sequence:
5'-ATTCTAGAGCGGCGCGGCGCATG-dt(30)BN-3' (where B = A,
C, or G and N = A, C, G, or T). Average insert size
1.35 kb (range 0.9-4.0 kb). 14/15 colonies contained
inserts by PCR. This library was enriched for full-length
clones and was constructed by Clontech Laboratories (Palo
Alto, CA)."

ORIGIN
Query Match 42.2%; Score 211.4; DB 5; Length 820;
Best Local Similarity 70.3%; Pred. No. 1.4e-30;
Matches 351; Conservative 0; Mismatches 121; Indels 27; Gaps 4;
QY 3 GGGGCTCACTCTGTGCCAGGCTGGAGTGGCGTGGATCTCAGATCACTACAACCT 62
DB 568 GAGTCTCACTCTGTGCCAGGCTGGAGTGGCGTGGATCTTGGCTCACTGCAAGCT 509
QY 63 CCATCTCTGGTTCAAATAATCTCTTGCCTCAGCCTCTCAAGTAGCTGGACTGTAG 122
DB 508 CCGCCTCTGGTTCAGTCAATCTCTCTGCTCAGCCTCTCAGTAGCTGGGACT----- 454
QY 123 CTCTCAAGTAGCTGGGACACACACACCATGCCAGCTAAATTTTGTGTTTTTTTGT 182
DB 453 -----ACAGGGCGCTGCCACTGGCGCGCTAAATTTTGT-----TATTTTACT 409
QY 183 AGACAGAGGTTTTCACCATGTTGGCCAGGCTGGGTGACCTCCCTTTTATGATTCCTCAT 242
DB 408 AGAGATGGGTTTTCACCATGTTAGCCAGGATGCTCGATCTCCTGAACTGTGATCCAC 349
QY 243 CTGCTCTATTCTTCCCTTTCTAATGACAGTATCCAGTTTCCTTATTCATTTATT 302
DB 348 CTCGCTC-----GGCCTCCCAAGTGTGGGATTACAGATGAGCCCGCGCCGCG 295
QY 303 ATATTCTTATTATTATGACAGAGTCTGTGCTTGTTCGCCAAGCTGGAGTACAGTGG 362
DB 294 CTTTTTTTTTTTTTTTGTGATGGAGTCTCATCTGTGCGCCAGGCTGGAGTGCAGTGG 235
QY 363 TGCATCTCGGCTCACTGCAAGTTCACCTGCTGGGTTACAGCATTTCTCCGCTCAGC 422
DB 234 GCGGATCTTGGCTCACTGCAAGTTCGCTCTCTGGGTTTCATGTCATTTCTCTGCTCAGC 175
QY 423 CTCCCCAGTAGCTGGGACTAAAGCGCTCGCCACCGCCCGCTAAATTTTGTATTT 482

Db 174 CTCCTGAGTAGCTGGGACTACAGCGCTGCCACTGCCCGCGCTAA-TTTTTGTATTT 116
QY 483 TTAATAAAGACGGGTTTC 501
DB 115 TTAGTAGAGATGGGTTTC 97
RESULT 6
CR627381 3775 bp mRNA linear HTC 22-SEP-2004
LOCUS Homo sapiens mRNA; cDNA DKFZp762F237 (from clone DKFZp762F237).
DEFINITION
CR627381
ACCESSION
CR627381.1 GI:50949857
VERSION
KEYWORDS
HTC.
SOURCE
Homo sapiens (human)
ORGANISM
Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini;
Hominidae; Homo.
REFERENCE
1 (bases 1 to 3775)
Ottewaelder,B.; Obermaier,B., Deutschenbaur,S., Schaipp,A.,
Mewes,H.W., Weil,B., Amid,C., Osanger,A., Fobo,G., Han,M. and
Wiemann,S.
The German cDNA Consortium
Direct Submission
Submitted (22-SEP-2004) MIPS, Ingolstaedter Landstr.1, D-85764
Neuherberg, GERMANY
Clone from S. Wiemann, Molecular Genome Analysis, German Cancer
Research Center (DKFZ); Email s.wiemann@dkfz-heidelberg.de;
sequenced by Medigenomix (Martinsried/Germany) within the cDNA
sequencing consortium of the German Genome Project.
This clone (DKFZp762F237) is available at the RZPD Deutsches
Ressourcenzentrum fuer Genomforschung GmbH in Berlin, Germany.
Please contact RZPD for ordering:
http://www.rzpd.de/cgi-bin/products/cl.cgi?cloneID=DKFZp762F237
Further information about the clone and the sequencing project is
available at http://mips.gsf.de/projects/cdna/.
FEATURES
Location/Qualifiers
1. .3775
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/clone="DKFZp762F237"
/tissue_type="melanoma (MeWo cell line)"
/clone_lib="762 (synonym: hm12). Vector pSport1; host
DH10B; sites NotI + SalI"
/dev_stage="adult"
/note="hypothetical protein, N-terminus truncated"
1. .3775
/gene="DKFZp762F237"
413. .744
/gene="DKFZp762F237"
/codon_start=1
/product="hypothetical protein"
/protein_id="CAH10478.1"
/db_xref="GI:50949858"
/translation="YGRNVVEALQEFQMOKSRGADLKNGLVYVEMVPSNPVVC
YVTFPGSCFGSFQFPTKABARRSAKIALMNSVFNHPRRIRIDEIKSVSALA
SPNGREADPNPTGIGAFREFMSENKGMLEFELMTVFQLLHWGSLKAMRERQC
SRQEVLAHYSHRALDDDIRHQWALDWVSRQSVFGALSRSLASTERLEARLAKEL
RFHKEKKDILVLAQLGNMHSNC"

ORIGIN
Query Match 42.0%; Score 210.6; DB 4; Length 3775;
Best Local Similarity 71.5%; Pred. No. 1.5e-30;
Matches 357; Conservative 0; Mismatches 89; Indels 53; Gaps 4;
QY 3 GGGGCTCACTCTGTGCCAGGCTGGAGTGGCGTGGATCTCAGATCACTACAACCT 62
DB 2219 GAGTCTTGTCTGTGCCAGGCTAGAGTGCAGTGGCGCATCTCGCTCACTGCAACT 2278
QY 63 CCATCTCTCGGTTCAAATAATCTCTTGCCTCAGCCTCTCAAGTAGCTGGGACTGTAG 122

```
Db 2279 CGCCTCCAGATTCACACCAATCTCTCGCTCAGCCTCCGAGTAGCTGGGA----- 2331
Qy 123 CTCTCAAGTAGCTGGCACACACCAATGCCAGCTAAATTTTGTGTGTTTTTGTGT 182
Db 2332 -----TTACAGGTGGCCACACACCAAGCTAAATTTTGTG-----TATTTTGT 2378
Qy 183 AGAGACAGGTTTTTCAACCATGTTGGCAGGCTGGTGACCTCCTTTAGATTCTCTCAT 242
Db 2379 AGAGACGGGGTTTCAACCATGTTAGCCAGGACGG-----TCTCCATCT 2420
Qy 243 CTGCTCTATTCTTCCCTCTTCTAATGCAAGTATCCAGTTTCTTACTTATCATCATTTATT 302
Db 2421 CTGACCTCGTATCCACCTGCTCGC-----CTCCCAACTATTGTG 2465
Qy 303 ATTATCTTATTATTATTGAGACAGAGTCTTGTCTTTGTGCCAAGGCTGGAGTACAGTGG 362
Db 2466 TGTTTTTTTTTTTTTTTTGGAGCGGAGTCTCGTCTGTGCCCGAGGCTGGAGTGG 2525
Qy 363 TCGATCTCGGCTACTGMACTGCACTGCTGGGTTTCAAGCCATCTCCCGCTCAGC 422
Db 2526 CGCTATCTGGCTACTGCACTGCACTGCGCTCCCGGTTTCAAGCCATCTCTCTGCTCAGC 2585
Qy 423 CTCCCCAGTAGCTGGGACTTAAAGCGCTTGCACACAGCCCGCGCTAAATTTTGTATT 482
Db 2586 CTCCTTAGTAGCTGGGACTTACAGCGCGCCGCGCACCATGCTGGCTAAATTTTGTATT 2645
Qy 483 TTAATAAAGACGGGGTTTC 501
Db 2646 TTAGTAGAGACGGGGTTTC 2664
```

```
RESULT 7
BZ605978
LOCUS WHAAT77TF Human MCF7 breast cancer cell line library (MCF7_1) Homo
DEFINITION sapiens genomic clone MCF7_1-5N10, genomic survey sequence.
ACCESSION BZ605978
VERSION BZ605978.1 GI:31514539
KEYWORDS GSS.
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
```

```
REFERENCE
AUTHORS Volik S., Zhao S., Chin K., Brebner J.H., Herndon D.R., Tao Q.,
Kowbel D., Huang G., Lapuk A., Kuo W.-L., Magrane G., de Jong P.,
Gray J.W. and Collins C.
End-sequence profiling: Sequence-based analysis of aberrant genomes
Proc. Natl. Acad. Sci. U.S.A. 100 (13), 7696-7701 (2003)
12788976
Contact: Volik SV
Colin Collins' lab
UCSF Comprehensive Cancer Center
UCSF Box 0808, San Francisco, CA 94143-0808, USA
Tel: 415 502 7066
Fax: 415 502 5665
Email: svolik@cc.ucsf.edu
This clone is available from Amplicon Express
http://www.genomex.com
Class: BAC ends.
```

```
FEATURES
source
1. 689
/organism="Homo sapiens"
/mol_type="genomic DNA"
/db_xref="taxon:9606"
/clone="MCF7_1-5N10"
/sex="female"
/clone_lib="Human MCF7 breast cancer cell line library
(MCF7_1)"
/notes="vector: pBCBAC1; Site 1: HindIII; This library was
constructed from MCF7 breast cancer cell line by Amplicon
Express (http://www.genomex.com) using their standard
```

```
ORIGIN
Query Match 41.9%; Score 209.8; DB 9; Length 689;
Best Local Similarity 67.1%; Pred. No. 3e-30;
Matches 335; Conservative 0; Mismatches 147; Indels 17; Gaps 2;
procedure."
Qy 3 GGGGCTCAGCTCTGTGCGCCAGGCTGGAGTGGGCGGTGATCTCAGATCACTACAACT 62
Db 89 GAGTCTTGCTCTGTGTTCCAGGCTGGAGTGCAGTGGTGTGATCTCGGCTCACTGCACT 148
Qy 63 CCATCTCTGCGTTCAAAATTAATTTCTTTCCTCAGCTCTCAAGTAGCTGGGACTTGTAG 122
Db 149 CTGCTCTCTGGGTTCAAAATGAATTTCTTTCCTCAGCTCTCCTGAGTAGCTGGGACTATAA- 207
Qy 123 CTCTCAAGTAGCTGCGACACACACCACTGCCAGCTAAATTTTGTGTTTTTGTGTTT 182
Db 208 -----GAGATGCCATCACACCCCTAAATTTTGTGTTTTTGTATTATT 253
Qy 183 AGAGACAGGTTTTTCAACCATGTTGGCAGGCTGGGCTGACCTCCCTTTTAGATTCTCCTCAT 242
Db 254 AGAGACAGGTTTTTCACTGTGTAAAGCAGGATGGTCTCAATCTCTGACCTCATGATCCAC 313
Qy 243 CTTGCTCTATTCTTTCCTTTCCTTCTTAATGCAAGTATCCAGTTTCCTTCTTATTCACATTTATT 302
Db 314 CCGCTCTGGCTCCCAAAATGCTGGGATTACAGGCGTGAGCCACCACTCTCTGGGCTTGT 373
Qy 303 ATTATCTTATTATTATTGAGACAGACTCTTGTCTTTCCTGCGCAAGGCTGGAGTACAGTGG 362
Db 374 TCTTTTTTTTTTTTTTTTGGAGCGAGTCTCACTTTATCGCCCGAGGCTGGAGTCTGTGG 433
Qy 363 TCGATCTCGGCTCAGTCAAGCTCCACCTGCTGGGTTTCAAGCATTTCTCCGCTCAGC 422
Db 434 CATATAATTTTGGTTCAGTGCACCTTCACTCTCTGGGTTTCAAGCAGTTCTCTGCTCAAC 493
Qy 423 CTCCCCAGTAGCTGGGACTAAAGCGCTTGCACACAGCCCGCTAAATTTTGTATT 482
Db 494 CTCCTGAGTAACCTGGGATTACAAGTGGCGCCACCACTGGTTAA- -TTTTGTATT 551
Qy 483 TTAATAAAGACGGGGTTTC 501
Db 552 TTAGTAGAGACGGGGTTTC 570
```

RESULT 8
BX496786/c
LOCUS DXFZp779J1329_r1 779 (synonym: hnccl) Homo sapiens cDNA clone
DEFINITION DXFZp779J1329 5', mRNA sequence.
ACCESSION BX496786
VERSION BX496786.1 GI:32012740
KEYWORDS EST.
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini;
Hominidae; Homo.
REFERENCE 1 (bases 1 to 642)
AUTHORS Ansoorge,W., Krieger,S., Regiert,T., Rittmueller,C., Schwager,B.,
Mewes,H.W., Weil,B., Amid,C., Osanger,A., Fobo,G., Han,M. and
Wiemann,S.
EST (Ansoorge,W., Krieger,S., Regiert,T., Rittmueller,C., et al.)
Unpublished (2003)
Contact: MIPS
MIPS
Ingolstaedter Landstr.1, D-85764 Neuherberg, Germany
This is the 5' sequence of the clone insert
Clone from S. Wiemann, Molecular Genome Analysis, German Cancer
Research Center (DKFZ); Email s.wiemann@dkfz-heidelberg.de;
sequenced by EMBL (European Molecular Biology Laboratories,
Heidelberg/Germany) within the cDNA sequencing consortium of the
German Genome Project.
No sl sequence available.


```

QY 483 TTAATAAGACGGGGTTTC 501
DB 537 TTAGTAGACGGGGTTTC 555

RESULT 10
CD723982/c
LOCUS
DEFINITION
oJ28g12.y1 Human lacrimal gland, unamplified: oJ Homo sapiens cDNA
clone oJ28g12 5', mRNA sequence.
ACCESSION
CD723982
VERSION
CD723982.1 GI:32274836
KEYWORDS
EST.
SOURCE
Homo sapiens (human)
ORGANISM
Homo sapiens
Eukaryota; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini;
Homidae; Homo.
REFERENCE
1 (bases 1 to 623)
Ozylidirim,A.M., Wistow,G.J., Gao,J., Wang,J., Dickinson,D.P.,
Frierson,H.F. Jr and Laurie,G.W.
The lacrimal gland transcriptome is an unusually rich source of
rare and poorly characterized gene transcripts
Invest. Ophthalmol. Vis. Sci. 46 (5), 1572-1580 (2005)
1581553
Contact: Wistow G
Section on Molecular Structure and Function
National Eye Institute
6/331, NIH, Bethesda, MD 20892-2740, USA
Tel: 301 402 3452
Fax: 301 496 0078
Email: graeme@helix.nih.gov
Plate: 28 row: 9 column: 12
Seq primer: M13Rpl reverse primer (ABI).
Location/Qualifiers
1..623
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/clone="oj28g12"
/tissue_type="lacrimal gland"
/dev_stage="Adult"
/lab_host="EMDH10B"
/clone_lib="Human lacrimal gland, unamplified: oj"
/notes="Organ: Eye; Vector: pCMVSPORT6; RNA was extracted
from 2 human lacrimal glands. A directionally cloned cDNA
library in the pCMVSPORT6 vector(Life Technologies) was
constructed at Bioserve Biotechnology (Laurel MD)
essentially following the protocols of the SuperScript
Plasmid System full details of which are contained in the
manufacturer's instruction manual
(http://www.lifetech.com/). First strand synthesis was
carried out using a Not I primer-adaptor
[5'-pGACTAGTTCAGCGCGGCC(C)(T)15-3']. EST analysis
was performed on the unamplified library at the NIH
Intramural Sequencing Center (NISC)."
```

ORIGIN
 Query Match 41.6%; Score 208.4; DB 6; Length 623;
 Best Local Similarity 68.8%; Pred. No. 5.6e-30;
 Matches 347; Conservative 0; Mismatches 146; Indels 11; Gaps 4;
 QY 3 GGGGCTCACTCTGTCGGCCAGCTGGAGTCGGGTGATCTCAGATCACTACACCT 62
 DB GAGTCTTGCTCTGTCGGCCAGCTGGAGTCGGGTGATCTCAGATCACTACACCT 439
 QY 63 CCATCTCTCTGGGTTCAATAATCTCTTCTGCTCAGCTCTCAAGTACCTGGGACTTGTAG 122
 DB CCGCCTCCAGGTTCAAGCAATCTCTGCTCAGCTCTCTGTCAGCTCTGTCAGTGGATTACAGG 379
 QY 123 CTCTCAAGTACCTGGCAGACACACACACACACACACACACACACACACACACACAC 182
 DB CAGGC-----GCCACACGCGCGCTAAATTTTTTTTTTTTTTTTTTTTTTTTGTATTTAGT 324

```

183 AGAGACAGGTTTTCACCATGTTGGCCAGGCTGG--GTGACCTCCCTTTTAGATTCTCCTC 240
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
323 AGAGACAGGATTTCCCATGTTGGCCAGGTTGGTCTCAATCTTCTGACCTCATGATCCAC 264
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
241 ATCTGTCTTATTTCTTCCCTTTCTTAATGCAAGTATCCAGTTTCTTACTTATCACAATTA 300
263 CCACCTTGGCTCCCAAGAGTGTGGGATTTACAGGACGCGCCACCATGCCCAGCTAATTTT 204
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
301 TTATTATCTTT--ATTATTATGACACAGATCTTGTCTTGTGCGCCAGGCTGGAGTAC 357
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
203 TGTITTTTGTGTGTTGTTTTCAGACGGGAGTTTCGCTCTGTGTCTCCAGGCTGGAATGC 144
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
358 AGTGGTGCATCTCGGCTCACTGCAAGCTCCACCTGCTGGGTTCAGGCAATTCCTCCGCC 417
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
143 AGTGGTGGATCTCAGTCTCACTGCAAGCTCGGCTCCGGGTTTCATGCCACTCTCTGCC 84
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
418 TCAGCTCCCGCAGTAGCTGGGACTTAAAGGCGCTTCGCCACACGCGCTAATTTTTTGG 477
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
83 TCAGCTCCCGCAGTAGCTGGGACTTACAGGCGCCACCACTGCTCGTCCGTTAA-TTTTTG 25
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
478 TATTTTAAATAAGACGGGGTTTC 501
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
24 TATTTTATGATATAGACGGGGTTTC 1
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

RESULT 11
BUI74155
LOCUS
DEFINITION
AGENCOURT 8106214 NIH_MGC_112 Homo sapiens cDNA clone IMAGE:6265610
5', mRNA sequence.
ACCESSION
BUI74155
VERSION
BUI74155.1 GI:22688139
KEYWORDS
EST.
SOURCE
Homo sapiens (human)
ORGANISM
Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini;
Homidae; Homo.
REFERENCE
1 (bases 1 to 891)
NIH-MGC http://mgi.nci.nih.gov/.
National Institutes of Health, Mammalian Gene Collection (MGC)
Unpublished (1999)
Contact: Robert Strausberg, Ph.D.
Email: rgs@bhs-xemail.nih.gov
Tissue Procurement: DCTD/DTP
cDNA Library Preparation: Rubin Laboratory
cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
DNA Sequencing by: Agencourt Bioscience Corporation
Clone distribution: MGC clone distribution information can be
found through the I.M.A.G.E. Consortium/LLNL at:
http://image.llnl.gov
Plate: LLCW2433 row: a column: 03
High quality sequence stop: 502.
Location/Qualifiers
1..891
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/clone="IMAGE:6265610"
/tissue_type="melanotic melanoma, cell line"
/lab_host="DH10B (phage-resistant)"
/clone_lib="NIH_MGC_112"
/notes="Organ: Skin; Vector: pOTB7; Site 1: XhoI; Site 2:
EcoRI; cDNA made by oligo-dT priming. Directionally cloned
into EcoRI/XhoI sites using the following 5' adaptor:
GGCAGCAG(G). Library constructed by Ling Hong in the
Laboratory of Gerald M. Rubin (University of California,
Berkeley) using ZAP-cDNA synthesis kit (Stratagene) and
Superscript II RT (Life Technologies). Note: this is a
NIH_MGC Library."
```

ORIGIN

source		1. 1654		/organism="Homo sapiens"	
				/mol_type="mRNA"	
				/db_xref="taxon:9606"	
				/clone="CS0DF001YF08"	
				/tissue_type="Fetal brain"	
				/plasmid="pCMVSPORT_6"	
ORIGIN					
Query Match		41.6%; Score 208.2; DB 4; Length 1654;			
Best Local Similarity		68.3%; Pred. No. 5.1e-30;			
Matches 341; Conservative		0; Mismatches 138; Indels 20; Gaps 3;			
QY	3	GGGGCTCACTCTGTCGCCAGGCTGGAGTGGGCGGTGATCTCAGATCACTACAACCT	62		
DB	420	GAGTCTCGCTCTGTCGCCAGGCTGGAGTGGGCGGTGATCTCAGATCACTACAACCT	479		
QY	63	CCATCTCTCGGTTCAAAATAATCTCTTGGCTCAGCCCTCTCAAGTAGCTGGGACTTTGTAG	122		
DB	480	CGGCTCTCGGTTTACGCCATCTCTCTGCTCAGCCCTCCCAAGTAGCTCGGACT-----	534		
QY	123	CTCTCAAGTAGCTGGCACACACCAATGCCAGCTAAATTTTTTGTGTGTTTTTTTGGT	182		
DB	535	-----ACAGGCGTCTCCACCATGCTGGCTTAATATT-----TGTATTTTTTAGT	580		
QY	183	AGAGACAGGTTTTCACCATGTTGGCCAGGCTGGGTGACCTCCCTTTTAGATTCTCCTCAT	242		
DB	581	AGAGACGGGGTTTACCGGTGTAGCCAGGATGGTCTCGATCTCTTAACCTGGTGATCCGC	640		
QY	243	CCTGCTCTATTCTTCCCTTTCTAATGCAATATCCAGTTTCTTACTTATCAATTTATT	302		
DB	641	CGGCTCGGCTCCCAAGTATTGGGATTACCGGCTGAGCCACCGCGCCGGCTGCAAA	700		
QY	303	ATTATTTCTATTATTATTGAGACAGAGTCTTGTGTTGCGCAAGGCTGGAGTACAGTGG	362		
DB	701	TAATCTTTCTTTTCTTGAGACAGAGTCTCGCTCTGTTGCCAGGCTGGAGTGCAGTGG	760		
QY	363	TGGATCTCGGCTCACTGCAAGCTCCACTGCTGGGTTTCAAGCATTTCTCCGCTCAGC	422		
DB	761	CAGATCTCGGCTACGGCAGCTCCGCTCCCGGTTTCAAGCATTTCTCTGCTCAGC	820		
QY	423	CTCCCCAGTAGCTGGGACTAAAGGCGCTGCCACACCGCCCGCTAAATTTTTTGTATTT	482		
DB	821	TTCCCGAGTAGCTGGGACTACAGGGGCCCGCACCGCCCGGCTAA-CTTTTGTGTTT	879		
QY	483	TTAATAAAGACGGGGTTTC 501			
DB	880	TTAGTAGAGACGGGGTTTC 898			
RESULT 13					
DR761502		777 bp mRNA linear EST 21-JUL-2005			
LOCUS		HESCA129_D09_g1_A037 NIH_MGC_262 Homo sapiens cDNA clone			
DEFINITION		IMAGE:7971211 5', mRNA sequence.			
ACCESSION		DR761502			
VERSION		DR761502.1 GI:71054202			
KEYWORDS		EST.			
SOURCE		Homo sapiens (human)			
ORGANISM		Homo sapiens			
REFERENCE					
AUTHORS		Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;			
TITLE		Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini;			
JOURNAL		Hominidae; Homo.			
REMARK		1 (bases 1 to 777)			
		NIH-MGC http://mgc.nci.nih.gov/.			
		National Institutes of Health, Mammalian Gene Collection (MGC)			
		Unpublished (1999)			
		Contact: Daniela S. Gerhard, Ph.D.			
		Office of Cancer Genomics			
		National Cancer Institute / NIH			
		Bldg. 31 Rm10A07 Bethesda, MD 20892			
		Email: cgabbs-r@mail.nih.gov			
		Tissue Procurement: BresaGen, Inc.			

Query Match		41.6%; Score 208.2; DB 5; Length 891;			
Best Local Similarity		68.3%; Pred. No. 5.7e-30;			
Matches 341; Conservative		0; Mismatches 138; Indels 20; Gaps 3;			
QY	3	GGGGCTCACTCTGTCGCCAGGCTGGAGTGGGCGGTGATCTCAGATCACTACAACCT	62		
DB	31	GAGTCTCGCTCTGTCGCCAGGCTGGAGTGGGCGGTGATCTCAGATCACTACAACCT	90		
QY	63	CCATCTCTCGGTTCAAAATAATCTCTTGGCTCAGCCCTCTCAAGTAGCTGGGACTTTGTAG	122		
DB	91	CCGCTCTCGGTTTACGCCATCTCTCTGCTCAGCTCCAGTCCCAAGTAGCTCGGACT-----	145		
QY	123	CTCTCAAGTAGCTGGCACACACCAATGCCAGCTAAATTTTTTGTGTGTTTTTTTGGT	182		
DB	146	-----ACAGGCGTCTCCACCATGCTGGCTTAATATT-----TGTATTTTTTAGT	191		
QY	183	AGAGACAGGTTTTCACCATGTTGGCCAGGCTGGGTGACCTCCCTTTTAGATTCTCCTCAT	242		
DB	192	AGAGACGGGGTTTACCGGTGTAGCCAGGATGGTCTCGATCTCTTAACCTGGTGATCCGC	251		
QY	243	CCTGCTCTATTCTTCCCTTTCTAATGCAATATCCAGTTTCTTACTTATCACATTTATT	302		
DB	252	CGGCTCGGCTCCCAAGTATTGGGATTACCGGCTGAGCCACCGCGCCGGCTGCAAA	311		
QY	303	ATTATTTCTATTATTATTGAGACAGAGTCTTGTGTTTGTGGCAAGGCTGGAGTACAGTGG	362		
DB	312	TAATCTTTCTTTTCTTGAGACAGAGTCTCGCTCTGTTGCCAGGCTGGAGTGCAGTGG	371		
QY	363	TGGATCTCGGCTCACTGCAAGCTCCACTGCTGGGTTTACGCCAATTTCTCCGCTCAGC	422		
DB	372	CACGATCTCGGCTACGGCAGCTCCGCTCCCGGTTTACGCCAATTTCTCTGCTCAGC	431		
QY	423	CTCCCCAGTAGCTGGGACTAAAGGCGCTGCCACACCGCCCGCTAAATTTTTTGTATTT	482		
DB	432	TTCCGAGTAGCTGGGACTACAGGGGCCCGCACCGCCCGGCTAA-CTTTTGTGTTT	490		
QY	483	TTAATAAAGACGGGGTTTC 501			
DB	491	TTAGTAGAGACGGGGTTTC 509			
CR591254		1654 bp mRNA linear HFC 21-JUL-2004			
LOCUS		full-length cDNA clone CS0DF001YF08 of Fetal brain of Homo sapiens			
DEFINITION		(human).			
ACCESSION		CR591254			
VERSION		CR591254.1 GI:50472061			
KEYWORDS		HFC; CNSLT CDNA.			
SOURCE		Homo sapiens (human)			
ORGANISM		Homo sapiens			
REFERENCE					
AUTHORS		Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;			
TITLE		Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini;			
JOURNAL		Hominidae; Homo.			
REMARK		1 (bases 1 to 1654)			
		Li, W. B., Gruber, C., Jessee, J. and Polayes, D.			
		Full-length cDNA libraries and normalization			
		Unpublished			
		Contact : Feng Liang Email : fliang@lifetech.com URL :			
		http://fulllength.invitrogen.com/ Invitrogen Corporation 1600			
		Faraday Avenue			
		2 (bases 1 to 1654)			
		Genoscope.			
		Direct Submission			
		Submitted (20-JUL-2004) Genoscope - Centre National de Sequencage :			
		BP 191 91006 EVRY cedex - FRANCE (E-mail : segret@genoscope.cns.fr			
		- Web : www.genoscope.cns.fr)			
		1st strand cDNA was primed with a NotI-oligo(dT) primer. Five prime			
		end enriched, double-strand cDNA was digested with Not I and cloned			
		into the Not I and EcoR V sites of the pCMVSPORT 6 vector. Library			
		was normalized. Library was constructed by Life Technologies, a			
		division of Invitrogen.			
		Location/Qualifiers			

FEATURES					
RESULT 12					
CR591254		1654 bp mRNA linear HFC 21-JUL-2004			
LOCUS		full-length cDNA clone CS0DF001YF08 of Fetal brain of Homo sapiens			
DEFINITION		(human).			
ACCESSION		CR591254			
VERSION		CR591254.1 GI:50472061			
KEYWORDS		HFC; CNSLT CDNA.			
SOURCE		Homo sapiens (human)			
ORGANISM		Homo sapiens			
REFERENCE					
AUTHORS		Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;			
TITLE		Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini;			
JOURNAL		Hominidae; Homo.			
COMMENT		1 (bases 1 to 1654)			
		Li, W. B., Gruber, C., Jessee, J. and Polayes, D.			
		Full-length cDNA libraries and normalization			
		Unpublished			
		Contact : Feng Liang Email : fliang@lifetech.com URL :			
		http://fulllength.invitrogen.com/ Invitrogen Corporation 1600			
		Faraday Avenue			
		2 (bases 1 to 1654)			
		Genoscope.			
		Direct Submission			
		Submitted (20-JUL-2004) Genoscope - Centre National de Sequencage :			
		BP 191 91006 EVRY cedex - FRANCE (E-mail : segret@genoscope.cns.fr			
		- Web : www.genoscope.cns.fr)			
		1st strand cDNA was primed with a NotI-oligo(dT) primer. Five prime			
		end enriched, double-strand cDNA was digested with Not I and cloned			
		into the Not I and EcoR V sites of the pCMVSPORT 6 vector. Library			
		was normalized. Library was constructed by Life Technologies, a			
		division of Invitrogen.			
		Location/Qualifiers			

cDNA Library Preparation: Express Genomics, Inc.
 cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
 DNA Sequencing by: Laboratory for Genomics and Bioinformatics,
 University of Georgia
 Clone distribution: MGC clone distribution information can be
 found through the I.M.A.G.E. Consortium/LLNL at:
 http://image.llnl.gov
 Plate: LLAM 17087 row: 9 column: 17
 Seq primer: JENREV (CAGGAAACAGCTATGACC)
 High quality sequence stop: 777.
 Location/Qualifiers
 1. 777
 /organism="Homo sapiens"
 /mol_type="mRNA"
 /db_xref="taxon:9606"
 /clone="IMAGE:7971211"
 /sex="male"
 /tissue_type="embryonic stem"
 /cell_type="human embryonic stem cells"
 /cell_lines="BG01"
 /lab_host="NIH MGC 262"
 /note="Vector: pExpress-1; Site 1: NotI; Site 2: EcoRV;
 RNA obtained from human embryonic stem cells isolated from
 the inner cell mass of blastocyst stage embryos and
 differentiated to an early neural progenitor cell type.
 Cell line id and NIH Registry designation is BG01.
 Positive for Nestin and Musashi expression. Passage number
 18. cDNA primed using oligo-dT primer:
 5'-pCAGTAGTTCTAGTCGCGCGGCCCT(7)25-3' and cloned into
 the EcoRV/NotI sites of pExpress-1. This primary library
 is non-normalized (normalized primary library is
 NIH MGC 259). It was constructed by Express Genomics
 (Frederick, MD). Sequence ends have been trimmed to
 exclude vector and regions below Phred quality 16. Note:
 this is a Mammalian Gene Collection library."

FEATURES
source

ORIGIN

Query Match 41.4%; Score 207.4; DB 8; Length 777;
 Best Local Similarity 69.9%; Pred. No. 8.4e-30;
 Matches 349; Conservative 0; Mismatches 96; Indels 54; Gaps 3;
 3 GGGGCTCACTCTGTGCGCCAGGCTGAGTGGGGTGGTGTGATCTCAGATCACTCAACCT 62
 Db 193 GAGTCTTCTCTGTGCGCCAGGCTAGAGTGCAGTGGCGGATCTCGGCTCACTCAACCT 252
 Qy 63 CCATCTCTGGTTCAAATAATCTCTTCTGCTCAGCCTCTCAAGTAGTGGGACTGTAG 122
 Db 253 CCGCTCCAGATTACACACCACTTCTCTGCTCAGCCTCCGAGTAGCTGGGA----- 305
 123 CTCTCAAGTAGCTGGCACACACCACTGCCAGCTAATTTTGTGTGTTTTTTTGGT 182
 Db 306 -----TTACAGGTGCGACCACTGCCAGCTAATTTTGTG-----TATTTTAGT 352
 Qy 183 AGAGACAGTTTTCACCATGTGGCCAGGCTGGGTGACCTCCCTTTTAGATTTCTCTCAT 242
 Db 353 AGAGACGGGTTTACCATGTAGCCAGGCG----- 385
 Qy 243 CTGCTCTATTTCTTCCCTTTCTAATGCAGTATCGAGTTTCTTACTTATCAATTTATT 302
 Db 386 -----TCTCCATCTCTGACCTCGTGATCACTGCTCGGCTCCCAAACTTATTTGT 438
 303 ATTATTTCTATTATTATTTGACAGAGTCTTGTCTTGTGCGCAAGGCTGGAGTACAGTGG 362
 Db 439 GTGTTTTTTTTTTTGTGACAGGAGTCTCGCTCTGTGCGCCAGGCTGGAGTGCAGTGG 498
 Qy 363 TGGCATCTCGGCTCACTGCAAGCTCCACCTGCTGGGTTTCAGCCATTTCTCCCGCTCAGC 422
 Db 499 CGCTATCTGGCTCACTGCAAGCTCCGCTCTCCGGTTTCAGCCACTCTCTGCTCAGC 558
 Qy 423 CTCCTAGTAGCTGGGACTAAGGGGCTGCGCACACGCGCCGCTAATTTTGTATT 482
 Db 559 CTCCTAGTAGCTGGGACTACAGGCGCGCGCACCATGCTGCTAATTTTGTATT 618

Qy 483 TTAATAAAGACGGGGTTTC 501
 Db 619 TTAGTAGAGACGGGGTTTC 637
 RESULT 14
 BUI80590 832 bp mRNA linear EST 04-SEP-2002
 LOCUS AGENCOURT_7852598 NIH_MGC_67 Homo sapiens cDNA clone IMAGE:6140102
 5', mRNA sequence.
 BUI80590
 BUI80590.1 GI:22694574
 EST.
 KEYWORDS
 SOURCE Homo sapiens (human)
 ORGANISM Homo sapiens
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini;
 Homnidae; Homo.
 1 (bases 1 to 832)
 NIH-MGC http://mgc.nci.nih.gov/.
 National Institutes of Health, Mammalian Gene Collection (MGC)
 Unpublished (1999)
 JOURNAL
 COMMENT Contact: Robert Strausberg, Ph.D.
 Email: cgapbs-re@mail.nih.gov
 Tissue Procurement: ATCC
 cDNA Library Preparation: Life Technologies, Inc.
 cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
 DNA Sequencing by: Agencourt Bioscience Corporation
 Clone distribution: MGC clone distribution information can be
 found through the I.M.A.G.E. Consortium/LLNL at:
 http://image.llnl.gov
 Plate: LLAM13458 Row: C column: 15
 High quality sequence stop: 508.
 Location/Qualifiers
 1. 832
 /organism="Homo sapiens"
 /mol_type="mRNA"
 /db_xref="taxon:9606"
 /clone="IMAGE:6140102"
 /tissue_type="retinoblastoma"
 /lab_host="DH10B (phage-resistant)"
 /clone_lib="NIH MGC 67"
 /note="Organ: eye; Vector: pCMV-SPORT6; Site 1: NotI;
 Site 2: Sall; Cloned unidirectionally. Primer: Oligo dT.
 Average insert size 1.75 kb. Library constructed by Life
 Technologies."

ORIGIN

Query Match 41.4%; Score 207.4; DB 5; Length 832;
 Best Local Similarity 69.9%; Pred. No. 8.3e-30;
 Matches 349; Conservative 0; Mismatches 96; Indels 54; Gaps 3;
 3 GGGGCTCACTCTGTGCGCCAGGCTGAGTGGGGTGGTGTGATCTCAGATCACTCAACCT 62
 Db 60 GAGTCTTGTCTGTGCGCCAGGCTAGAGTGCAGTGGCGGATCTCGGCTCACTCAACCT 119
 Qy 63 CCATCTCTGGTTCAAATAATTTCTCTGCTCAGCCTCTCAAGTAGTGGGACTGTAG 122
 Db 120 CCGCTCCAGATTACACCACTTCTCTGCTCAGCCTCCGAGTAGCTGGGA----- 172
 123 CTCTCAAGTAGCTGGCACACACCACTGCCAGCTAATTTTGTGTGTTTTTTTGGT 182
 Db 173 -----TTACAGGTGCGACCACTGCCAGGCTAGCCAGGACGG----- 252
 Qy 183 AGAGACAGTTTTCACCATGTGGCCAGGCTGGGTGACCTCCCTTTTAGATTTCTCTCAT 242
 Db 220 AGAGACGGGTTTACCATGTAGCCAGGCTAGCCAGGACGG----- 252
 Qy 243 CCTGCTCTATTCTTCCCTTTCTAATGCAGTATTCAGTTTCTTACTTATCAATTTATT 302
 Db 253 -----TCTCCATCTCTGACCTCGTGATCCACCTCGGCTCCCAAACTTATTTGT 305

Qy 303 ATTATCTTATTATTAGACAGAGCTCTGCTTGTGCGCAAGGCTGGAGTACAGTGG 362

Db 306 GTGTTTTTTTTTTTTTTTGGACGAGTCTCGCTCTGTGCGCCAGGCTGGAGTGCAGTGG 365

Qy 363 TGGCATCTCGGCTCAGTCAAGCTCCACCTGCTGGGTTACGCCATTCTCCGGCTCAGC 422

Db 366 CGCTATCTCGGCTCACTCAAGCTCCGCTCCGGGTTACGCCATCTCTCGCTCAGC 425

Qy 423 CTCCCCAGTAGCTGGGACTAAAGGCGCTGCCACACAGCCCGCTAAATTTTTTTGTAATT 482

Db 426 CTCCTTAGTAGCTGGGACTACAGGCGCGCCACCATGCGCTGCTAAATTTTTTTGTAATT 485

Qy 483 TTAATAAAGACGGGGTTTC 501

Db 486 TTAGTAGAGACGGGGTTTC 504

RESULT 15

BUB38401

LOCUS

DEFINITION

ACCESSION

VERSION

KEYWORDS

SOURCE

ORGANISM

REFERENCE

AUTHORS

TITLE

JOURNAL

COMMENT

1 (bases 1 to 935)

NIH-MGC <http://mgc.nci.nih.gov/>.

National Institutes of Health, Mammalian Gene Collection (MGC)

Unpublished (1999)

Contact: Robert Strausberg, Ph.D.

Email: cgapbs-r@mail.nih.gov

Tissue Procurement: Dr. James R. Lupski

cDNA Library Preparation: Life Technologies, Inc.

cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)

DNA Sequencing by: Agencourt Bioscience Corporation

Clone distribution: MGC clone distribution information can be found through the I.M.A.G.E. Consortium/LLNL at: <http://image.llnl.gov>

Plate: LLAM13554 row: e column: 03

High quality sequence stop: 622.

Location/Qualifiers

1..935

/organism="Homo sapiens"

/mol_type="mRNA"

/db_xref="taxon:9606"

/clone="IMAGE:6177002"

/sex="male"

/tissue_type="dorsal root ganglia"

/dev_stage="adult, 36 yr"

/lab_host="DH10B"

/clone_lib="Lupski dorsal root ganglion"

/note="Vector: pCMV-SPORT6 (Life Technologies); Site 1: Not1; Site 2: SalI; cDNA made by oligo-dT priming. Directionally cloned using the following adaptors: 5'-GACTAGTTTAGTCGAGCGGCCGCTT(15)-3'. Size selected > 1 kb for average insert length 1.7 kb. This is a primary library, non-amplified. Library constructed by Life Technologies and donated by J. Lupski, M.D./Ph.D. (Baylor College of Medicine) and is available through Life Technologies."

ORIGIN

Query Match 41.4%; Score 207.4; DB 5; Length 935;

Best Local Similarity 69.9%; Pred. No. 8.1e-30;

Matches 349; Conservative 0; Mismatches 96; Indels 54; Gaps 3;

Qy 3 GGGGCTCACTCTCTGCGCCAGGCTGGAGTGGCGGTGATCTCAGATCACTACACCT 62

Db 247 GAGTCTGTCTCTGCGCCAGGCTAGAGTGCAGTGGCGCATCTCGGCTCACTGCAACCT 306

Qy 63 CCATCTCCTGGGTTCAAAATAATTCTCTTGGCTCAGCTCTCAAGTAGCTGGACTTTAG 122

Db 307 CCGCTCCAGATTTCACACCAATTTCTCTGCTCAGCTCCGAGTAGCTGGGA----- 359

Qy 123 CTCTCAAGTAGCTGGGACACACCAATGCCAGTAAATTTTTTGTGTGTTTTTTGGT 182

Db 360 -----TTACAGGTGCCACACACGCCAGCTAAATTTTTTG-----TAATTTTAGT 406

Qy 183 AGAGACAGGTTTTTCAACCATGTGGCCAGGCTGGGTGACCTCCCTTTTATAGATTCCTCAT 242

Db 407 AGAGCGGGTTTCAACCATGTAGCCAGGACGG----- 439

Qy 243 CTGCTCTATTCTTCCCTTTCTAATGCAGTATCCAGTTTCTTACTTATCACATTATT 302

Db 440 -----TCTCCATCTCTGACCTCGTGATCCACCTGCGCTCCCAAACTATTGT 492

Qy 303 ATATTTCTTATTATTATTTAGACAGAGTCTTGTCTTGTGCGCAAGGCTGGAGTACAGTGG 362

Db 493 GTGTTTTTTTTTTTTTTTGGACGAGTCTCGCTCTGTGCGCCAGGCTGGAGTGCAGTGG 552

Qy 363 TGGCATCTCGGCTCAGTCAAGCTCCACCTGCTGGGTTACGCCATTCTCCCGCTCAGC 422

Db 553 CGCTATCTCGGCTCAGTCAAGCTCCGCTCCGGGTTACGCCACTCTCCTGCTCAGC 612

Qy 423 CTCCCCAGTAGCTGGGACTAAAGGCGCTGCCACACAGCCCGCTAAATTTTTTTGTAATT 482

Db 613 CTCCTTAGTAGCTGGGACTACAGGCGCGCCACCATGCGCTGCTGCTGCTGCTGCTGCT 672

Qy 483 TTAATAAAGACGGGGTTTC 501

Db 673 TTAGTAGAGACGGGGTTTC 691

Search completed: February 9, 2006, 18:10:14

Job time : 2367.65 secs

GenCore version 5.1.7
Copyright (c) 1993 - 2006 Bioacceleration Ltd.

OM nucleic - nucleic search, using sw model

Run on: February 9, 2006, 13:28:05 ; Search time 10772.2 Seconds
(without alignments)
10559.003 Million cell updates/sec

Title: US-10-607-806-1_COPY_7200_9200

Perfect score: 1998.6

Sequence: i ctactcagctgtccctccc.....ckctacataaatatctatta 2001

Scoring table: IDENTITY_NUC

Gapop 10.0 , Gapext 1.0

Searched: 5883141 seqs, 28421725653 residues

Total number of hits satisfying chosen parameters: 11766282

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : GenEmbl.*

1: gb_ba.*

2: gb_in.*

3: gb_env.*

4: gb_on.*

5: gb_ov.*

6: gb_pat.*

7: gb_ph.*

8: gb_pt.*

9: gb_ro.*

10: gb_sts.*

11: gb_sy.*

12: gb_un.*

13: gb_vi.*

14: gb_htg.*

15: gb_pl.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB	ID	Description
1	1998.6	100.0	8368	8	AY438977	Homo sapi
2	1998.6	100.0	13612	6	AX377239	Sequence
C 3	1998.6	100.0	122302	8	AC003982	Homo sapi
C 4	1998.6	100.0	220384	14	AC078926	Homo sapi
5	1811.2	90.6	189729	14	AC073930	Homo sapi
C 6	462.6	23.1	178887	14	AC068981	Homo sapi
7	462.6	23.1	190162	8	AC009244	Homo sapi
8	460.6	23.0	123192	14	AP000589	Homo sapi
9	460.4	23.0	150266	8	AC022408	Homo sapi
10	459.4	23.0	167996	8	AC021753	Homo sapi
11	457.4	22.9	137693	8	AC003689	Homo sapi
12	456.2	22.8	201460	8	AP003721	Homo sapi
13	451.2	22.6	195986	14	AC116933	Papio anu
C 14	449.2	22.5	134210	8	AC005052	Homo sapi
C 15	447.8	22.4	131943	8	AC005484	Homo sapi
C 16	447.8	22.4	136119	8	HS1028D15	Human DNA
17	447.8	22.4	186197	14	AL627384	Homo sapi
18	446.2	22.3	163681	8	AL136992	Human DNA

19	445	22.3	117336	8	AC093171	Homo sapi
20	442.6	22.1	91927	8	AC004771	Homo sapi
C 21	441.8	22.1	200430	8	AC011500	Homo sapi
C 22	439.4	22.0	180076	14	AC022177	Homo sapi
C 23	438.8	22.0	201227	14	AC160565	Pan trogl
C 24	438.6	21.9	163704	14	AC141415	Pan trogl
C 25	438.2	21.9	199045	14	AC149081	Pan trogl
C 26	437.8	21.9	185574	14	AC161822	Pan trogl
C 27	435.8	21.8	203790	8	AC010422	Homo sapi
C 28	435	21.8	119043	8	AC139451	Homo sapi
C 29	435	21.8	163542	14	AC129071	Pan trogl
C 30	434.8	21.8	96995	8	AC008149	Homo sapi
C 31	434.2	21.7	152659	8	AL591503	Human DNA
C 32	433.8	21.7	19384	8	AP003531	Homo sapi
C 33	433.2	21.7	89448	8	AC005067	Homo sapi
C 34	433.2	21.7	178976	14	AL331095	Homo sapi
C 35	432.6	21.6	156933	8	AC020750	Homo sapi
C 36	432.2	21.6	81117	8	AC016898	Homo sapi
C 37	432.2	21.6	192961	8	AP002770	Homo sapi
C 38	432	21.6	153392	8	AL133548	Human DNA
C 39	432	21.6	235183	14	AC160951	Colobus g
C 40	431.4	21.6	214031	14	AC160952	Colobus g
C 41	431	21.6	134957	8	HSJ930L11	Human DNA
C 42	430.8	21.6	44062	8	AC004517	Homo sapi
C 43	430.8	21.6	118396	8	AC073138	Homo sapi
C 44	430.8	21.6	162701	14	AC073317	Homo sapi
C 45	430.6	21.5	41613	8	AC004790	Homo sapi

ALIGNMENTS

RESULT 1	AY438977	8368 bp	DNA	linear	PRI 29-OCT-2003
LOCUS	Homo sapiens phospholipase A2, group IB (pancreas) (PLA2G1B) gene, complete cds.				
DEFINITION	AY438977				
ACCESSION	AY438977.1 GI:37953284				
VERSION	Homo sapiens (human)				
KEYWORDS	Eukaryota; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini; Hominidae; Homo.				
SOURCE	1 (bases 1 to 8368)				
ORGANISM	Rieder, M.J., Livingston, R.J., Daniels, M.R., Chung, M.-W., Miyamoto, K.B., Nguyen, C.P., Nguyen, D.A., Poel, C.L., Robertson, P.D., Schackwitz, W.S., Sherwood, J.K., Witrak, L.A. and Nickerson, D.A.				
REFERENCE	Direct Submission				
AUTHORS	Submitted (16-OCT-2003) Genome Sciences, University of Washington, 1705 NE Pacific, Seattle, WA 98195, USA				
TITLE	To cite this work please use: NIEHS-SNPs, Environmental Genome Project, NIEHS ES15478, Department of Genome Sciences, Seattle, WA				
JOURNAL	(URL: http://egp.gs.washington.edu)				
COMMENT					
FEATURES	Location/Qualifiers				
source	1..8368				
repeat_region	/organism="Homo sapiens"				
variation	/mol_type="genomic DNA"				
variation	/db_xref="taxon:9606"				
repeat_region	101..257				
variation	/rpt_family="MIR"				
variation	/rpt_type="dispersed"				
repeat_region	205				
variation	/frequency="0.01"				
variation	/replace="a"				
repeat_region	306				
variation	/frequency="0.01"				
variation	/replace="a"				
repeat_region	326..538				
variation	/rpt_family="Alu"				
misc_feature	/rpt_type="dispersed"				
misc_feature	354..496				


```
/gene="PLA2G1B"
/frequency="0.01"
/replace="a"
5949
variation
/gene="PLA2G1B"
/frequency="0.01"
/replace="g"
6039
variation
/gene="PLA2G1B"
/frequency="0.01"
/replace="a"
6082
variation
/gene="PLA2G1B"
/frequency="0.14"
/replace="t"

Query Match
Best Local Similarity 99.7%; Pred. No. 0; DB 8; Length 8368;
Matches 1995; Conservative 6; Mismatches 0; Indels 0; Gaps 0;

QY 1 CTATCTCAGCTGTCCTCCCACTTTCCAGGTGCTGCCAGACACATGCAACAATGCTAYGAC 60
Db 3574 CTATCTCAGCTGTCCTCCCACTTTCCAGGTGCTGCCAGACACATGCAACAATGCTATGAC 3633
QY 61 CAGGCCAAGAGCTGGACAGCTGTAAATTTCTGCTGGACAMMCCGTACACCCACACCTAT 120
Db 3634 CAGGCCAAGAGCTGGACAGCTGTAAATTTCTGCTGGACAAACCCGTACACCCACACCTAT 3693
QY 121 TCATACTCCTGCTCGGCTCGGCAATCACTGTAGCAGTGGTTTATCCCTTCTCTTGACC 180
Db 3694 TCATACTCCTGCTCGGCTCGGCAATCACTGTAGCAGTGGTTTATCCCTTCTCTTGACC 3753
QY 181 TATGAATTTAGTTGGTTCTCAGTAGGCGGGGGGAATAATAGTAACAACAGCCATGAT 240
Db 3754 TATGAATTTAGTTGGTTCTCAGTAGGCGGGGGGAATAATAGTAACAACAGCCATGAT 3813
QY 241 TTAGTGTTAAATTTCTGGTCTCGGCGAGTGTCTCTTTAATCTCTCAGAAACAACATATG 300
Db 3814 TTAGTGTTAAATTTCTGGTCTCGGCGAGTGTCTCTTTAATCTCTCAGAAACAACATATG 3873
QY 301 GGATAGGTACAAATATCTCTTAACAGATAAGAAACTGAGGCTCAGAGGCTGAGCT 360
Db 3874 GGATAGGTACAAATATCTCTTAACAGATAAGAAACTGAGGCTCAGAGGCTGAGCT 3933
QY 361 ATTTGCCCAAGATCACACAGCTGTGAAGTGGTGACAGTTTGGGTTTTTTTTTTTGTGTGT 420
Db 3934 ATTTGCCCAAGATCACACAGCTGTGAAGTGGTGACAGTTTGGGTTTTTTTTTTTGTGTGT 3993
QY 421 TTAGAGACAGGCTCTTGCTCTGTCAACCAGGATGAGCACAGTGGTGCAACCATAGGTCA 480
Db 3994 TTAGAGACAGGCTCTTGCTCTGTCAACCAGGATGAGCACAGTGGTGCAACCATAGGTCA 4053
QY 481 CTGAGGCTCAACCTCTGAGCTCAGGGATCTGCTGACCTCAGCTCCCAAGTAGCTGG 540
Db 4054 CTGAGGCTCAACCTCTGAGCTCAGGGATCTGCTGACCTCAGCTCCCAAGTAGCTGG 4113
QY 541 GACTACGAGCTGTGACACACAGCTGCTTAATTAATAAATAATTTTTTTGTAGAGACTGGG 600
Db 4114 GACTACGAGCTGTGACACACAGCTGCTTAATTAATAAATAATTTTTTTGTAGAGACTGGG 4173
QY 601 TCTTACTAGTTGGCAGGCTGTCTTAACTCTCTGGCTTCAAGCAATCTCTCTACCTTG 660
Db 4174 TCTTACTAGTTGGCAGGCTGTCTTAACTCTCTGGCTTCAAGCAATCTCTCTACCTTG 4233
QY 661 GCATCCCAAGTGTGGGATTTACAGGGTGAACCAACATGCGGCTACTTATTTCTTTA 720
Db 4234 GCATCCCAAGTGTGGGATTTACAGGGTGAACCAACATGCGGCTACTTATTTCTTTA 4293
QY 721 CATTTCCATCTTTTCCAAATAGATGTAAATCCACAGAACAGGGATTTACTGCTATTTCTT 780
Db 4294 CATTTCCATCTTTTCCAAATAGATGTAAATCCACAGAACAGGGATTTACTGCTATTTCTT 4353
QY 781 CCTTTCTTTTTTGAGACAGAGTCTCACTTCATCACTCAACCTCCGCTTCAGCTCACTGCA 840
```

```
Db 4354 CCTTTCTTTTTTGACACAGAGTCTCACTTCATCACTCAACCTCGCTCAGCTCACTGCA 4413
QY 841 ACCTCTGCTCCCGGGTTCAAGYGAATCTCTCTGCTAGCTAGCTCTGCTAGCTAGCTGAATTA 900
Db 4414 ACCTCTGCTCCCGGGTTCAAGCGAATCTCTCTGCTAAGCTCTCTGCTAGCTAGCTGAATTA 4473
QY 901 CAAGCGTCACACCAATGCTTTGGCTAAATTTTTTTTATTTTTTTAGCAGAGATGGGGTTTTTAC 960
Db 4474 CAAGCGTCACCAATGCTTTGGCTAAATTTTTTTTATTTTTTTAGCAGAGATGGGGTTTTTAC 4533
QY 961 CATGTTGCCCAGGCTGGTCTCAAACTCTCTGACCTCAAGTGAATCTGCTGCTCACTGCTCC 1020
Db 4534 CATGTTGCCCAGGCTGGTCTCAAACTCTCTGACCTCAAGTGAATCTGCTGCTCACTGCTCC 4593
QY 1021 CAAGTGTCTGGAATTTATAGGCGTGAGTCACTGCTGCTGGCGATTACTGTCTATTTCTT 1080
Db 4594 CAAGTGTCTGGAATTTATAGGCGTGAGTCACTGCTGCTGGCGATTACTGTCTATTTCTT 4653
QY 1081 TATTGCTATATCCCAAGATCTAGAGCAGTGTCTGACATATATAGTAGTGTCTCAATAATAA 1140
Db 4654 TATTGCTATATCCCAAGATCTAGAGCAGTGTCTGACATATATAGTAGTGTCTCAATAATAA 4713
QY 1141 TTGATGAATGACACAGCCTAGATATAAACTTTCTTTTTTTTAAACAACTCTTGACA 1200
Db 4714 TTGATGAATGACACAGCCTAGATATAAACTTTCTTTTTTTTAAACAACTCTTGACA 4773
QY 1201 ACTTTGACAGATAAATAAATCTTGTCACTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1260
Db 4774 ACTTTGACAGATAAATAAATCTTGTCACTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 4833
QY 1261 TTTTTCATATGCTCTCAAACTTTATGTTACTGTTTTTCACTGTTTTTCACTGTTTTTATGCTAC 1320
Db 4834 TTTTTCATATGCTCTCAAACTTTATGTTACTGTTTTTCACTGTTTTTCACTGTTTTTATGCTAC 4893
QY 1321 TGAATAATATGCTCTTAAATTTGCTTATACATCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 1380
Db 4894 TGAATAATATGCTCTTAAATTTGCTTATACATCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 4953
QY 1381 ACAATCTGATGAAGCTATGAACCTCTCTCCAGAGAAATACACACACACACACACT 1440
Db 4954 ACAATCTGATGAAGCTATGAACCTCTCTCCAGAGAAATACACACACACACACACT 5013
QY 1441 CACACACAGTTTTTTTTTAAATTTGCTTAACTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 1500
Db 5014 CACACACAGTTTTTTTTTAAATTTGCTTAACTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 5073
QY 1501 TGTTCATATTAATTAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATA 1560
Db 5074 TGTTCATATTAATTAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATA 5133
QY 1561 ACTTTGGAAGTCCAGGCTGGTGGATCACTTTGAGGTGAGAGTTTGGAGACAGGCTGGTCT 1620
Db 5134 ACTTTGGAAGTCCAGGCTGGTGGATCACTTTGAGGTGAGAGTTTGGAGACAGGCTGGTCT 5193
QY 1621 AATATGTTGAACCTCTCTACTATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATA 1680
Db 5194 AATATGTTGAACCTCTCTACTATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATA 5253
QY 1681 GCCTGTAGTCCAGCTACTCTCGGAGGCTGAGGCAAGAGAAATGCTTGAACCTTGGGAGGCA 1740
Db 5254 GCCTGTAGTCCAGCTACTCTCGGAGGCTGAGGCAAGAGAAATGCTTGAACCTTGGGAGGCA 5313
QY 1741 GAGGTTGAGTGGAGGAGATCCCACTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1800
Db 5314 GAGGTTGAGTGGAGGAGATCCCACTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 5373
QY 1801 CTATCTCAAAAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATA 1860
Db 5374 CTATCTCAAAAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATA 5433
QY 1861 TCCTGAAGGTAAAGCAGAGATACGTAATATATATATATATATATATATATATATATATATATATAT 1920
```

Db	5434	TCCTGAAGGTAAAGCAGAGATACGTAATAATTATATGTATTAAGAGTTTAAATGCAATTTTAACT	5493
Qy	1921	GTAATCTTATTTGTTTATTTTGGTTATAAAAAGTAAACCAAGCCAAAAGTAATGCAACTTCAA	1980
Db	5494	GTAATCTTATTTGTTTATTTTGGTTATAAAAAGTAAACCAAGCCAAAAGTAATGCAACTTCAA	5553
Qy	1981	ACKCTACATAAAATATCTATTA 2001	
Db	5554	ACTCTACATAAAATATCTATTA 5574	
RESULT 2			
LOCUS	AX377239	13612 bp	DNA linear PAT 18-MAR-2002
DEFINITION	Sequence 1 from Patent WO0212562.		
ACCESSION	AX377239		
VERSION	AX377239.1	GI:19573528	
KEYWORDS	Homo sapiens (human)		
SOURCE	Homo sapiens		
ORGANISM	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini; Hominidae; Homo.		
REFERENCE	1		
AUTHORS	Kazemi, A., Kliem, S.E. and Koshy, B.		
TITLE	Haplotypes of the pla2g1b gene		
JOURNAL	Patent: WO 0212562-A 1 14-FEB-2002;		
	Genasense Pharmaceuticals, Inc. (US)		
FEATURES	Location/Qualifiers		
source	1..13612		
	/organism="Homo sapiens"		
	/mol_type="unassigned DNA"		
	/db_xref="taxon:9606"		
variation	3845		
	/note="PS1: polymorphic base G or A"		
variation	3968		
	/note="PS2: polymorphic base A or G"		
variation	6060		
	/note="PS2: polymorphic base A or G"		
variation	6844		
	/note="PS4: polymorphic base G or A"		
variation	9531		
	/note="PS5: polymorphic base G or A"		
ORIGIN			
Query Match 100.0%; Score 1998.6; DB 6; Length 13612;			
Best Local Similarity 99.8%; Pred. No. 0;			
Matches 1996; Conservative 5; Mismatches 0; Indels 0; Gaps 0;			
Qy	1	CTATCTCAGCTGTCCTCCCACTTTCCAGGTGCTGCCAGACATGACAACTGCTAYGAC	60
Db	6716	CTATCTCAGCTGTCCTCCCACTTTCCAGGTGCTGCCAGACATGACAACTGCTATGAC	6775
Qy	61	CAGGCCAAGAGCTGGACAGCTGTAATTTCTGCTGGACAMMCCGTACACCCACACCTAT	120
Db	6776	CAGGCCAAGAGCTGGACAGCTGTAATTTCTGCTGGACAMMCCGTACACCCACACCTAT	6835
Qy	121	TCATATCTCTGCTCTGGCTCGGCAATCACCTGTAGCAGTAGGTTTATCCCTTCCTTGACC	180
Db	6836	TCATATCTCTGCTCTGGCTCGGCAATCACCTGTAGCAGTAGGTTTATCCCTTCCTTGACC	6895
Qy	181	TATGAATTCATGTTGGTTCTCAGTAGCGCGGGGAAATAATAGTAACACAGCCATGAT	240
Db	6896	TATGAATTCATGTTGGTTCTCAGTAGCGCGGGGAAATAATAGTAACACAGCCATGAT	6955
Qy	241	TTAGTGTAAATTTTCTTGGTCTGGCAGTGCTCTCTTAACTCCTCAGAACACACTATG	300
Db	6956	TTAGTGTAAATTTTCTTGGTCTGGCAGTGCTCTCTTAACTCCTCAGAACACACTATG	7015
Qy	301	GGATAGGTACAAATTAATCTCTCACTTAAACAGATAAGAAAACCTGAGGCTCAGAAAGCTGAGCT	360
Db	7016	GGATAGGTACAAATTAATCTCTCACTTAAACAGATAAGAAAACCTGAGGCTCAGAAAGCTGAGCT	7075

Qy	361	ATTTGCCCAAGATCACACAGCTTTGTAAGTGTGACAGTTTGGGTTTTTTTTTGTGTGTGT	420
Db	7076	ATTTGCCCAAGATCACACAGCTTTGTAAGTGTGACAGTTTGGGTTTTTTTTTGTGTGTGT	7135
Qy	421	TTAGAGACAGGGTCTTGCTCTGTCAACCAGGCATGAGCACAGTGGTGCAACCATAGGTCA	480
Db	7136	TTAGAGACAGGGTCTTGCTCTGTCAACCAGGCATGAGCACAGTGGTGCAACCATAGGTCA	7195
Qy	481	CTGAGGCTCAACCTCTGAGCTCAAGGGATCTGCTGACCTCAGCCCTCCCAAGTAGCTGG	540
Db	7196	CTGAGGCTCAACCTCTGAGCTCAAGGGATCTGCTGACCTCAGCCCTCCCAAGTAGCTGG	7255
Qy	541	GACTACGAGGTGCACCAACACCGCTCGCTAAATTAATAAATAATTTTTGTAGACACTGGG	600
Db	7256	GACTACGAGGTGCACCAACACCGCTCGCTAAATTAATAAATAATTTTTGTAGACACTGGG	7315
Qy	601	TCTTACTACGTTGGCCAGGCTTGTTCTTAAACTCTCTGCTTCAAGCAATCTCTACCTTG	660
Db	7316	TCTTACTACGTTGGCCAGGCTTGTTCTTAAACTCTCTGCTTCAAGCAATCTCTACCTTG	7375
Qy	661	GCATCCCAAGTGTCTGGATTACAGGGGTGAGCCACCATGTGCGGTACTTATTTCTTTA	720
Db	7376	GCATCCCAAGTGTCTGGATTACAGGGGTGAGCCACCATGTGCGGTACTTATTTCTTTA	7435
Qy	721	CATTCCATCTTTCCAAATAGAAATGAAGATCCACAGAACAGGGAATTAATCTTCTT	780
Db	7436	CATTCCATCTTTCCAAATAGAAATGAAGATCCACAGAACAGGGAATTAATCTTCTT	7495
Qy	781	CCTTCTCTTTTGTAGACAGAGTCTCATTCAATCACTCAACCTCGGTTCACTGCA	840
Db	7496	CCTTCTCTTTTGTAGACAGAGTCTCATTCAATCACTCAACCTCGGTTCACTGCA	7555
Qy	841	ACCTCTGCTCCCGGTTCAAGYGAATCTCTGCTGAAGCCCTCTGAGTACTGGAATTA	900
Db	7556	ACCTCTGCTCCCGGTTCAAGYGAATCTCTGCTGAAGCCCTCTGAGTACTGGAATTA	7615
Qy	901	CAAGGTGCACACCATGCTTGGCTAAATTTTGTATTATTTAGCAGAGATGGGGTTTTAC	960
Db	7616	CAAGGTGCACACCATGCTTGGCTAAATTTTGTATTATTTAGCAGAGATGGGGTTTTAC	7675
Qy	961	CATGTTGCCAGGCTGGTCTCAAACTCCTGACCTCAAGTGAATCTGCTGCTCAGTCTCC	1020
Db	7676	CATGTTGCCAGGCTGGTCTCAAACTCCTGACCTCAAGTGAATCTGCTGCTCAGTCTCC	7735
Qy	1021	CAAGTGTGGAAATTAAGGCGTGAGTCACTGTGCTGGCGGATTAATCTTATTTCTT	1080
Db	7736	CAAGTGTGGAAATTAAGGCGTGAGTCACTGTGCTGGCGGATTAATCTTATTTCTT	7795
Qy	1081	TATTGCTATATCCCCAGATCTAGACAGTGTCTGACATATAGTAGGTGCTCAATAAATAA	1140
Db	7796	TATTGCTATATCCCCAGATCTAGACAGTGTCTGACATATAGTAGGTGCTCAATAAATAA	7855
Qy	1141	TGTGAATATGACAGCCTAGATATAAACTTTCTTTTCTTTTAAAAACAATCTTGACA	1200
Db	7856	TGTGAATATGACAGCCTAGATATAAACTTTCTTTTCTTTTAAAAACAATCTTGACA	7915
Qy	1201	ACTTTGCGAATAAATAAATCTTGCAATCTGCTTTTCACTTATCATCCTGTTATGACT	1260
Db	7916	ACTTTGCGAATAAATAAATCTTGCAATCTGCTTTTCACTTATCATCCTGTTATGACT	7975
Qy	1261	TTTTTCATATGCTCAAAACCTTTATTTGTTACTGTTTTTCAATTTACTATTTTAGTCAC	1320
Db	7976	TTTTTCATATGCTCAAAACCTTTATTTGTTACTGTTTTTCAATTTACTATTTTAGTCAC	8035
Qy	1321	TGAATAATATGCTTTAAATTTGCTTATACATCTCTGCTCCAATTTAGAGGCCAAATTT	1380
Db	8036	TGAATAATATGCTTTAAATTTGCTTATACATCTCTGCTCCAATTTAGAGGCCAAATTT	8095
Qy	1381	ACAATCTGATGAAGCTATGAACCTCTCCCCAGAGAAATACACACACACACACT	1440
Db	8096	ACAATCTGATGAAGCTATGAACCTCTCCCCAGAGAAATACACACACACACACT	8155
Qy	1441	CACACACAGTTTTTTTTTAAATGTTTGCACATAAGACAAGAAAACCTGCAATTAGAGGATGTT	1500


```
Db      8156  |||||CACACACAGT|||TTTAAATGTTTTCACAACTAAGACAAGANACCTGCATTAGAGATGTT 8215
Qy      1501  TGTTCATATTAATAAATAAATACTCAGTTGGGCACAGTGACTCAAGCCCTGTAAACACACAGT 1560
Db      8216  TGTTCATATTAATAAATAAATACTCAGTTGGGCACAGTGACTCAAGCCCTGTAAACACACAGT 8275
Qy      1561  ACTTTGGAGTCCAAAGGTGGGTGGATCACTCAGGTGAGAAATTCGAGACCAAGCTGGTC 1620
Db      8276  ACTTTGGAGTCCAAAGGTGGGTGGATCACTCAGGTGAGAAATTCGAGACCAAGCTGGTC 8335
Qy      1621  AATATGGTCGAACCCCTATCTCTACTAAAATAACAAAATTTAGCTGGGTCTAGTGATGCAT 1680
Db      8336  AATATGGTGAAACCCCTATCTCTACTAAAATAACAAAATTTAGCTGGGTCTAGTGATGCAT 8395
Qy      1681  GCCTGTAGTCCAGCTACTCGGGAGCTGAGGCAAGAGAAATTCCTTGAACCTGGGAGGCA 1740
Db      8396  GCCTGTAGTCCAGCTACTCGGGAGCTGAGGCAAGAGAAATTCCTTGAACCTGGGAGGCA 8455
Qy      1741  GAGGTTCAGTGAGCGCGAGATCCACCACTGCATCCAGCCTGGCGCACACAGCGAGACT 1800
Db      8456  GAGGTTCAGTGAGCGCGAGATCCACCACTGCATCCAGCCTGGCGCACACAGCGAGACT 8515
Qy      1801  CTATCTCAAAAATAAATAAATAAATAAAGATCGAGAGAAACAAACCTAATAAGAT 1860
Db      8516  CTATCTCAAAAATAAATAAATAAATAAAGATCGAGAGAAACAAACCTAATAAGAT 8575
Qy      1861  TCCTCAAGTGAAGCAGAGATAGTAAATTTATATGTAATAAAGTTTAAATGATTTTAACT 1920
Db      8576  TCCTCAAGTGAAGCAGAGATAGTAAATTTATATGTAATAAAGTTTAAATGATTTTAACT 8635
Qy      1921  GTAATCTTATGTTTATTTTGGTTATAAAGTAAACAAAGCCAAAGTAACTCAACTCAA 1980
Db      8636  GTAATCTTATGTTTATTTGGTTATAAAGTAAACAAAGCCAAAGTAACTCAACTCAA 8695
Qy      1981  ACKTACATAAATATCTATTA 2001
Db      8696  ACTCTACATAAATATCTATTA 8716

RESULT 3
AC003982/c
LOCUS      Homo sapiens PAC clone 166H1 from 12q, complete sequence. PRI 18-MAR-1999
DEFINITION
ACCESSION  AC003982
VERSION    AC003982.1 GI:2769695
KEYWORDS   HTG.
SOURCE     Homo sapiens (human)
ORGANISM   Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini;
            Hominidae; Homo.
REFERENCE  1 (bases 1 to 122302)
            Sulston, J.E. and Waterston, R.
            Toward a complete human genome sequence
            Genome Res. 8 (11), 1097-1108 (1998)
            9847074
REFERENCE  2 (bases 1 to 122302)
            Bradshaw, H., Wu, X. and Ozersky, P.
            The sequence of Homo sapiens PAC clone 166H1
            Unpublished (1999)
REFERENCE  3 (bases 1 to 122302)
            Waterston, R.
            Direct Submission
            Submitted (13-JAN-1998) Department of Genetics, Washington
            University, 4444 Forest Park Avenue, St. Louis, Missouri 63108, USA
            4 (bases 1 to 122302)
            Waterston, R.
            Direct Submission
            Submitted (18-MAR-1999) Department of Genetics, Washington
            University, 4444 Forest Park Avenue, St. Louis, Missouri 63108, USA
            SUBMITTED BY: WUGSC
            Genome Sequencing Center
```

Department of Genetics
Washington University
St. Louis MO 63108, USA
<http://genome.wustl.edu/gac>
<mailto:Sapiens@watson.wustl.edu>

NOTICE: This sequence may not represent the entire insert of this clone. It may be shorter because we only sequence overlapping clone sections once, or longer because we provide a small overlap between neighboring data submissions.

This sequence was finished as follows unless otherwise noted: all regions were double stranded, sequenced with an alternate chemistry, or covered by high quality data (i.e., phred quality >= 30); an attempt was made to resolve all sequencing problems, such as compressions and repeats; all regions were covered by sequence from more than one subclone; and the assembly was confirmed by restriction digest.

MAPPING INFORMATION:

This clone was originally isolated in the laboratory of Professor Graeme Bell, Howard Hughes Medical Institute and Departments of Biochemistry and Molecular Biology, and Medicine, The University of Chicago, Chicago, IL, USA. The clone was provided by the laboratory of Dr. Roger Cox at The Wellcome Trust Centre For Human Genetics, Oxford, UK. Some contig information was also obtained from Yamagata et al., Nature 384:455-8 (1996).

SOURCE INFORMATION:

This clone was derived from human PAC library RPCI-1, prepared by Pieter de Jong and coworkers at the Roswell Park Cancer Institute (<http://bacpac.med.buffalo.edu>) using the method described by Ioannou et al., Nature Genetics 6:84-9 (1994). The library is from one male donor.

The clone may be obtained either from Genome Systems, Inc.

(<http://www.genomesystems.com>) or Research Genetics, Inc.

(<http://www.resgen.com>); or from Pieter de Jong.

VECTOR: pCYPAC2

NEIGHBORING SEQUENCE INFORMATION:

The clone sequenced to the left is 278C19; the clone sequenced to the right is 15E1. Actual start of this clone is at base position 1 of 166H1; actual end is at 122302 of 166H1.

FEATURES	Location/Qualifiers
source	1..122302
	/organism="Homo sapiens"
	/mol_type="genomic DNA"
	/db_xref="taxon:9606"
	/chromosome="12"
	/map="12q"
	/clone="166H1"
	/clone_lib="RPCI-1"
repeat_region	1..238
	/rpt_family="Alu"
repeat_region	239..265
	/rpt_family="AT-rich"
repeat_region	268..565
	/rpt_family="Alu"
repeat_region	1001..1309
	/rpt_family="Alu"
repeat_region	1440..1745
	/rpt_family="Alu"
repeat_region	1909..1940
	/rpt_family="MER1_type"
repeat_region	1955..1983
	/rpt_family="AT-rich"
repeat_region	2063..2362
	/rpt_family="Alu"
repeat_region	2516..2650
	/rpt_family="Alu"
repeat_region	2674..2730
	/rpt_family="L1"
repeat_region	2735..2788
	/rpt_family="(GAAA)n"

```
repeat_region 2790..3077
/rpt_family="Alu"
repeat_region 3081..3139
/rpt_family="Li"
repeat_region 3716..4020
/rpt_family="Alu"
repeat_region 4196..4492
/rpt_family="Alu"
gene 4560..14034
/genes="WUGSC:H 166H1.1"
CDS join(4560..5056,13418..13748,13882..14034)
/genes="WUGSC:H 166H1.1"
/notes="unknown_function; 60% similar to Z50177
(PID:g927403) (PID:g927402); H_166H1.1"
/codon_start=1
/evidence=not experimental
/protein_id="AAB95634.1"
/db_xref="GI:2769696"
/translation="MKMSFALTFRSAKGRWIANPSQPCSKASIGLFVPASPPLDPEKV
KELQRFITLSKLLVMTGAGISTESGIPDYSEKVLGYARTDRRPIQHGDVFRSAPIR
QRYWARNFVGNPQFSSHQPNPAHWALSTWEKLGKLYLWLTQNVDAHPTKAGSRRLTEL
HGMEDRAVCSVSLGSRVCLDCEGTGPRGLQERFQVLPNTSAAEAGLAPGDVDF
LSEEOVRGFQVPTQCQCGHLKPDVVFEGDTVPDKVDFVHKRVKEADSLLVVGSLLQ
VYSGYRFTLTAEKKLPAILNIGPTRSDDLACLKLSRCGELLPLIDPC"
repeat_region 5366..5655
/rpt_family="Alu"
repeat_region 5672..5798
/rpt_family="Alu"
repeat_region 5802..5834
/rpt_family="(TA)n"
repeat_region 5834..5857
/rpt_family="POLY_A"
repeat_region 5870..6184
/rpt_family="Alu"
repeat_region 6224..6268
/rpt_family="(CA)n"
repeat_region 6269..6558
/rpt_family="Alu"
repeat_region 6729..7029
/rpt_family="Alu"
repeat_region 7051..7172
/rpt_family="Alu"
repeat_region 7192..7500
/rpt_family="Alu"
repeat_region 7518..7820
/rpt_family="Alu"
repeat_region 8207..8500
/rpt_family="Alu"
repeat_region 8576..8878
/rpt_family="Alu"
repeat_region 9023..9154
/rpt_family="Alu"
repeat_region 9165..9470
/rpt_family="Alu"
repeat_region 9887..10188
/rpt_family="Alu"
repeat_region 10221..10342
/rpt_family="L1"
repeat_region 10408..10707
/rpt_family="Alu"
repeat_region 10713..11014
/rpt_family="Alu"
repeat_region 11020..11110
/rpt_family="Alu"
repeat_region 11146..11386
/rpt_family="Alu"
repeat_region 11208..11386
/rpt_family="Alu"
repeat_region 11391..11515
/rpt_family="Alu"
repeat_region 11568..11705
/rpt_family="Alu"
repeat_region 11737..12032
```

```

/rpt_family="Alu"
12034..12066
/rpt_family="7SLRNA"
12075..12372
/rpt_family="Alu"
12448..12753
/rpt_family="Alu"
12763..12874
/rpt_family="L2"
12999..13303
/rpt_family="Alu"
/notes="complement(13674..13752)
complement(13840..14195)
/notes="match to EST N80020 (NID:g1242721) za91a08.s1"
/notes="match to EST N80020 (NID:g1242721) za91a08.s1"
14196..14229
/rpt_family="AT-rich"
14240..14339
/rpt_family="U6"
14349..14646
/rpt_family="Alu"
14652..14940
/rpt_family="Alu"

Query Match 100.0%; Score 1998.6; DB 8; Length 122302;
Best Local Similarity 99.7%; Pred. No. 0;
Matches 1995; Conservative 6; Mismatches 0; Indels 0; Gaps 0;

QY 1 CTATCTCAGCTGTCCCTCCCACTTTTCCAGGTGCTGCCAGACACATGACAACTGCTATGAC 60
Db 26088 CTATCTCAGCTGTCCCTCCCACTTTTCCAGGTGCTGCCAGACACATGACAACTGCTATGAC 26029

QY 61 CAGGCCAAGAAGCTGGACAGCTGTAAATTTCTGCTGGCAAMMCGTACACCCACACCTAT 120
Db 26028 CAGGCCAAGAAGCTGGACAGCTGTAAATTTCTGCTGGCAAMMCGTACACCCACACCTAT 25969

QY 121 TCATACCTCTGCTCTGGCTCGGCAATCACCTGTAGCAGTAGGTTTATCCCTTCCTTGACC 180
Db 25968 TCATACCTCTGCTCTGGCTCGGCAATCACCTGTAGCAGTAGGTTTATCCCTTCCTTGACC 25909

QY 181 TATGAATTTCTAGTTGGTTCTCAGTAGGCCGGGGGAAATATAGTAAACAAGCCATGAT 240
Db 25908 TATGAATTTCTAGTTGGTTCTCAGTAGGCCGGGGGAAATATAGTAAACAAGCCATGAT 25849

QY 241 TTAGTGTAAATTTTCTGCTGCGGAGTCTCTCTTTAACTCCCTCAGAACAACTATG 300
Db 25848 TTAGTGTAAATTTTCTGCTGCGGAGTCTCTCTTTAACTCCCTCAGAACAACTATG 25789

QY 301 GGATAGGTACAAATATCTCACTTAAACAGATAAGAAACTCAGGCTCAGAGGCTGAGCT 360
Db 25788 GGATAGGTACAAATATCTCACTTAAACAGATAAGAAACTCAGGCTCAGAGGCTGAGCT 25729

QY 361 ATTTGCCCAAGATCACACAGCTTGTAAAGTGTGACAGTTGGGTTTTTTTTTTTGTGTGT 420
Db 25728 ATTTGCCCAAGATCACACAGCTTGTAAAGTGTGACAGTTGGGTTTTTTTTTTTGTGTGT 25669

QY 421 TTAGAGACAGGGTCTTGTCTGTACCCAGGCAATGAGCAGCTGGTGCACCACTAGTCA 480
Db 25668 TTAGAGACAGGGTCTTGTCTGTACCCAGGCAATGAGCAGCTGGTGCACCACTAGTCA 25609

QY 481 CTGACGCTCAACCTCTGAGCTCAAGGATCTGTGACCTCAGCCTCCCAAGTAGCTGG 540
Db 25608 CTGACGCTCAACCTCTGAGCTCAAGGATCTGTGACCTCAGCCTCCCAAGTAGCTGG 25549

QY 541 GACTACGAGCTGCACCACTCAGCCTGGCTAAATTAATAAAAAATTTTTTTGTAGAGCTGG 600
Db 25548 GACTACGAGCTGCACCACTCAGCCTGGCTAAATTAATAAAAAATTTTTTTGTAGAGCTGG 25489

QY 601 TCTTACTAGTTGGCCAGGCTTGTCTTAAACTCCCTGGCTTCAAGCAATCTCTCTACTCTG 660
Db 25488 TCTTACTAGTTGGCCAGGCTTGTCTTAAACTCCCTGGCTTCAAGCAATCTCTCTACTCTG 25429

QY 661 GCATCCCAAAAGTGTGGGATTACAGGGGTGAGCCACCATGTGCGGCTACTTATTTCTTTA 720
```

Db 25428 GATCCCAAGAGCTGGGATTAACAGGGGTAGCCACCATGTCGGCTACTTATTTCTTTA 25369
QY 721 CATTCCTATCTTCCAAATAGATGTAAGATCCACGAGACAGGATTTACTGCTTATTTCTT 780
Db 25368 CATTCCTATCTTCCAAATAGATGTAAGATCCACGAGACAGGATTTACTGCTTATTTCTT 25309
QY 781 CCTTTCTTTTGTGACAGAGTCTCACTTCATCATCTCAACCTCGGTTGAGTCTCACTGCA 840
Db 25308 CCTTTCTTTTGTGACAGAGTCTCACTTCATCATCTCAACCTCGGTTGAGTCTCACTGCA 25249
QY 841 ACCTTCGCTCCGGGTTCAAGYGATTTCTTCGCTCAAGCTTAAGCTCTGAGTAGCTGAATTA 900
Db 25248 ACCTTCGCTCCGGGTTCAAGCGATTTCTTCGCTCAAGCTCTGAGTAGCTGAATTA 25189
QY 901 CAAGCGTGACACCATCTGCTGGCTAAATTTTGTGATTTTGTAGCAGAGATGGGTTTTTAC 960
Db 25188 CAAGCGTGACACCATCTGCTGGCTAAATTTTGTGATTTTGTAGCAGAGATGGGTTTTTAC 25129
QY 961 CATGTTGCCAGGCTGGTCTCAAACTCCTGACCTCAAGTGATCTGCTGCCCTCAGTCTCC 1020
Db 25128 CATGTTGCCAGGCTGGTCTCAAACTCCTGACCTCAAGTGATCTGCTGCCCTCAGTCTCC 25069
QY 1021 CAAGTGCTGGAATTTATAGGCTGAGTCACTGCTGCCGATTAAGTCTTATTTCTT 1080
Db 25068 CAAGTGCTGGAATTTATAGGCTGAGTCACTGCTGCCGATTAAGTCTTATTTCTT 25009
QY 1081 TATTGCTATATCCAGATCTAGAGCAGTGTCTGACATATAGTAGTGTCTCAATAAATAA 1140
Db 25008 TATTGCTATATCCAGATCTAGAGCAGTGTCTGACATATAGTAGTGTCTCAATAAATAA 24949
QY 1141 TTGATGATGACAGCCTAGATATAAACTTTCTTTCTTTTCTTTTAAACAACTCTGACA 1200
Db 24948 TTGATGATGACAGCCTAGATATAAACTTTCTTTCTTTTAAACAACTCTGACA 24889
QY 1201 ACTTTGAGATAAATAAATCTTGCAATCTGCTTTTTCATATACCTTTGATGACT 1260
Db 24888 ACTTTGAGATAAATAAATCTTGCAATCTGCTTTTTCATATACCTTTGATGACT 24829
QY 1261 TTTTCATATGCTTCAACCTTTATGTTAGTGTCTTTTCTGCTTCTTCTTCTTCTTCTTCTT 1320
Db 24828 TTTTCATATGCTTCAACCTTTATGTTAGTGTCTTTTCTGCTTCTTCTTCTTCTTCTTCTT 24769
QY 1321 TGAATAATATGCTTAAATTTGCTTTATATACATCTCTGCTGCACCTTTAGAGGCCAAATTT 1380
Db 24768 TGAATAATATGCTTAAATTTGCTTTATATACATCTCTGCTGCACCTTTAGAGGCCAAATTT 24709
QY 1381 ACAATCTGATGAAAGCTATGAACCTCTCCCGAGAGAAATACACACACACACACT 1440
Db 24708 ACAATCTGATGAAAGCTATGAACCTCTCCCGAGAGAAATACACACACACACACT 24649
QY 1441 CACACACAGTTTTTTTTTAAATGTTGCACTAAGCAAGAACCTGCAATAGAGGATGTT 1500
Db 24648 CACACACAGTTTTTTTTTAAATGTTGCACTAAGCAAGAACCTGCAATAGAGGATGTT 24589
QY 1501 TGTTTATATTAATTAATAAATCACTGCTGGGCAAGTCAAGCTGTAACCAAGT 1560
Db 24588 TGTTTATATTAATTAATAAATCACTGCTGGGCAAGTCAAGCTGTAACCAAGT 24529
QY 1561 ACTTTGAAAGTCCAAGTGGGTGATCACTTGAGGTGAGAAGTTGAGACCAAGCTGGTC 1620
Db 24528 ACTTTGAAAGTCCAAGTGGGTGATCACTTGAGGTGAGAAGTTGAGACCAAGCTGGTC 24469
QY 1621 AATATGTAAGACCTTATCTACTAATAAATAAATAAATGCTGGGTGAGTGTGATGTCAT 1680
Db 24468 AATATGTAAGACCTTATCTACTAATAAATAAATAAATGCTGGGTGAGTGTGATGTCAT 24409
QY 1681 GCCTGTAGTCCAGCTACTCGGAGGCTGAGGCAAGAGAAATGCTTCAACCTGGAGGCA 1740
Db 24408 GCCTGTAGTCCAGCTACTCGGAGGCTGAGGCAAGAGAAATGCTTCAACCTGGAGGCA 24349
QY 1741 GAGGTTGAGTGAGCGAGATCCACCACTGCACTCCAGCTGGGCGACACAGCGAGACT 1800

Db 24348 GAGGTTGAGTGAGCGAGATCCACCACTGCACTCCAGCTGGGCGACACAGCGAGACT 24289
QY 1801 CTATCTCAAAAAATAAATAAATAAAGATCGAGAGAACAAAATAAATAGAT 1860
Db 24288 CTATCTCAAAAAATAAATAAATAAAGATCGAGAGAACAAAATAAATAGAT 24229
QY 1861 TCCTGAAGGTAAGCAGAGATACGTAATATATATGTAATAAAGTTTAAATGCAATTTAACT 1920
Db 24228 TCCTGAAGGTAAGCAGAGATACGTAATATATATGTAATAAAGTTTAAATGCAATTTAACT 24169
QY 1921 GTAATCTTATTTTATTTTGGTTATATAAAGTAACAAAGCAAAAGTAATGCAACTTCAA 1980
Db 24168 GTAATCTTATTTTATTTTGGTTATATAAAGTAACAAAGCAAAAGTAATGCAACTTCAA 24109
QY 1981 ACKTACATATAATCTATTATTA 2001
Db 24108 ACTTACATAAATATCTATTATTA 24088
AC078926 220384 bp DNA linear HTG 26-MAR-2002
Homo sapiens chromosome 12 clone RP11-836M11, WORKING DRAFT
SEQUENCE, 11 unordered pieces.
AC078926
AC078926.23 GI:19718703
HTG: HTGS PHASE1; HTGS_DRAFT; HTGS_FULLTOP.
KEYWORDS
Homo sapiens (human)
ORGANISM
Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini;
Hominidae; Homo.
REFERENCE
1 (bases 1 to 220384)
Muzny, D.M., Adams, C., Adio-Oduola, B., Ali-osman, F.R., Allen, C.,
Alsbrooks, S.L., Amarantunga, H.C., Are, J.R., Ayele, M., Banks, T.,
Barbieri, J., Benton, J., Bimige, K., Blankenburg, K., Bonnin, D.,
Bouck, J., Bowie, S., Brieva, M., Brown, E., Brown, M., Bryant, N.P.,
Buhay, C., Burch, P., Burkett, C., Burrell, K.L., Byrd, N.C.,
Carroll, T.F., Carter, M., Cavazos, S.R., Chacko, J., Chavez, D.,
Chen, G., Chen, R., Chen, Z., Chowdhry, I., Christopoulos, C.,
Cleveland, C.D., Cox, C., Coyle, M.D., Dathorne, S.R., David, R.,
Davila, M.L., Davis, C., Davy-Carroll, L., Dederich, D.A.,
Dellaway, K.R., Delgado, O., Denn, A.L., Ding, Y., Dinh, H.H.,
Douthwaite, K.J., Draper, H., Dugan-Rocha, S., Durbin, K.J.,
Earnhart, C., Edgar, D., Edwards, C.C., Elhaj, C., Escotto, M.,
Fallis, T., Ferraguto, D., Flagg, N., Ford, J., Foster, P., Frantz, P.,
Gabisi, A., Gao, J., Garcia, A., Garner, T., Garza, N., Gill, R.,
Gorrell, J.H., Guevara, W., Gunaratne, P., Hale, S., Hamilton, K.,
Harris, C., Harris, K., Hart, M., Havlak, P., Hawes, A., Hernandez, J.,
Hernandez, O., Hodgson, A., Hogues, M., Holloway, C., Hollins, B.E.,
Homs, F., Howard, S., Huber, J., Hulyk, S., Hume, J., Jackson, L.E.,
Jacobson, B., Jia, X., Johnson, R., Jolivet, S., Joudan, S.,
Karleson, E., Kelly, S., Khan, U., King, L., Korvah, J., Kovar, C.,
Kratovic, J., Kureshi, A., Landry, N., Leal, B., Lewis, L.C., Lewis, L.,
Li, J., Li, Z., Lichtarge, O., Lieu, C., Liu, J., Liu, W., Ljulesevic, H.,
Lozano, R.J., Lu, X., Lucier, A., Lucier, R., Luna, R., Ma, J.,
Maheehari, M., Mapua, P., Martin, R., Martindale, A., Martinez, E.,
Massey, E., Mawhney, E., McLeod, M.P., Meador, M., Mei, G., Metzker, M.,
Miner, G., Miner, Z., Mitchell, T., Mohabbat, K., Morgan, M., Morris, S.,
Moser, M., Neal, D., Newton, J., Newton, N., Nguyen, A., Nguyen, N.,
Nguyen, N., Nickerson, E., Nwokenkwo, S., Ogih, M., Okwuonu, G.,
Oragunye, N., Oviedo, R., Pace, A., Payton, B., Peery, J., Perez, L.,
Peters, L., Pickens, R., Primus, E., Pu, L.L., Quiles, M., Ren, Y.,
Rives, M., Rojas, A., Rojibokan, I., Rolfe, M., Ruiz, S., Savery, G.,
Scherrer, S., Scott, G., Shen, H., Shoshitari, N., Sisson, I.,
Sodergren, E., Sonaik, T., Sparks, A., Stanley, H., Stone, H.,
Sutton, A., Svatek, A., Taber, P., Tamerisa, A., Tamerisa, K., Tang, H.,
Tansey, J., Taylor, C., Taylor, T., Telford, B., Thomas, N., Thomas, S.,
Usmani, K., Vasquez, L., Vera, V., Villalon, D., Vinson, R., Wang, Q.,
Wang, S., Ward-Moore, S., Warren, R., Washington, C., Watlington, S.,
Williams, G., Williamson, A., Wlezyk, R., Woodson, S., Worley, K.,
Wu, C., Wu, Y., Zhou, J., Zorrilla, S., Nelson, D.,
Weinstock, G. and Gibbs, R.

TITLE	Direct Submission	gap	2157. .2256	
JOURNAL	Unpublished	gap	/estimated_length=unknown	
REFERENCE	2 (bases 1 to 220384)	gap	7113. .7212	
AUTHORS	Worley, K.C.	gap	/estimated_length=unknown	
TITLE	Direct Submission	gap	13284. .13383	
JOURNAL	Submitted (11-AUG-2000) Human Genome Sequencing Center, Department of Molecular and Human Genetics, Baylor College of Medicine, One Baylor Plaza, Houston, TX 77030, USA	gap	/estimated_length=unknown	
	3 (bases 1 to 220384)	gap	20661. .20760	
REFERENCE	Worley, K.C.	gap	/estimated_length=unknown	
AUTHORS	Direct Submission	gap	34474. .34573	
TITLE	Submitted (26-MAR-2002) Human Genome Sequencing Center, Department of Molecular and Human Genetics, Baylor College of Medicine, One Baylor Plaza, Houston, TX 77030, USA	gap	/estimated_length=unknown	
JOURNAL	On Mar 26, 2002 this sequence version replaced gi:18449664.	gap	44395. .44494	
COMMENT	----- Genome Center of Medicine	gap	57127. .57226	
	Center code: BCM	gap	/estimated_length=unknown	
	Web site: http://www.hgsc.bcm.tmc.edu/	gap	80267. .80366	
	Contact: hgsc-help@bcm.tmc.edu	gap	/estimated_length=unknown	
	----- Project Information	gap	109399. .109498	
	Center project name: HCM	gap	/estimated_length=unknown	
	Center clone name: RP11-836M11	gap	149813. .149912	
	----- Summary Statistics		/estimated_length=unknown	
	Sequencing vector: Plasmid;			
	Chemistry: Dye-terminator Big Dye; 47% of reads			
	Assembly program: Phrap; version 0.990329			
	Consensus quality: 234783 bases at least Q40			
	Consensus quality: 241821 bases at least Q30			
	Consensus quality: 245671 bases at least Q20			
	Estimated insert size: 219187; sum-of-contigs estimation			
	Quality coverage: 8.9x in Q20 bases; sum-of-contigs estimation			

	* NOTE: Estimated insert size may differ from sequence length			
	(see http://www.hgsc.bcm.tmc.edu/docs/Genbank_draft_data.html).			
	* NOTE: This is a 'working draft' sequence. It currently			
	* consists of 11 contigs. The true order of the pieces			
	* is not known and their order in this sequence record is			
	* arbitrary. Gaps between the contigs are represented as			
	* runs of N, but the exact sizes of the gaps are unknown.			
	* This record will be updated with the finished sequence.			
	* as soon as it is available and the accession number will			
	* be preserved.			

	1 2156: contig of 2156 bp in length			
	2157 2256: gap of unknown length			
	2257 7112: contig of 4856 bp in length			
	7113 7212: gap of unknown length			
	7213 13283: contig of 6071 bp in length			
	13284 13383: gap of unknown length			
	13384 20660: contig of 7277 bp in length			
	20661 20760: gap of unknown length			
	20761 34473: contig of 13713 bp in length			
	34474 34573: gap of unknown length			
	34574 44394: contig of 9821 bp in length			
	44395 44495: gap of unknown length			
	44495 57126: contig of 12632 bp in length			
	57127 57226: gap of unknown length			
	57227 80266: contig of 23040 bp in length			
	80267 80366: gap of unknown length			
	80367 109398: contig of 29032 bp in length			
	109399 109498: gap of unknown length			
	109499 149812: contig of 40314 bp in length			
	149813 149912: gap of unknown length			
	149913 220384: contig of 70472 bp in length.			
FEATURES	Location/Qualifiers			
source	1. .220384			
	/organism="Homo sapiens"			
	/mol_type="genomic DNA"			
	/db_xref="taxon:9606"			
	/chromosome="12"			
	/clone="RP11-836M11"			

Db 90323 GCCTAACCTATTAAATGTAATTTGATGTTGGCATATAAAATCCCTAGCCAGTTTGT 90264
Qy 760 GGGATTACTGCTATTTCTCTCTTTTGTGAGACAGAGCTCTCACTTCATCACTCA 819
Db 90263 TTTGTT-TTGTGTTTGTGAGACAGAGTTTGTCTGTGGACTGCACATGAGTGCAGTG 90205
Qy 820 ACCTCCGTTAGCTCACTGCAACCTCTGCTCCCGGTTCAAGYGAATCTCTGCTCAAG 879
Db 90204 GCTCAATCTCGGCTCACTGCAACCTCCGCTCCCGGTTCAAGCGAATCTCTGCTCAG 90145
Qy 880 CTTCTGAGTACGTAATCAAGCTGCGGACCATGCTTGGCTAATTTTGTATTT 939
Db 90144 CTTCTGAGTACGTAATCAAGCTGCGGACCATGCTTGGCTAATTTTGTATTT 90085
Qy 940 TTAGCAGAGATGGGCTTTTACCATGTTGCCAGGCTGTTCTCAAACTCTCACTCAAGT 999
Db 90084 TTGTTAGAGTCAAGTTTAAACCAATGTTGCCAGGCTGTTTCAAACTCTCACTCAAGT 90025
Qy 1000 GATGCTGCTGCTCACTCTCCAAAGTCTGGAATTTAGGCTGAGTCACTGTGCTGG 1059
Db 90024 GATCCACCTGCTCACTCTCCAAAGTCAAGGATTTACAGGCTGAGCCACGCCGCGG 89965
Qy 1060 CCGATTACTGCTATTTTCTTTTATGCTATATCCCACTAGAGAGTGTCTGACATA 1119
Db 89964 CCAACAACTTTTCAATTAAGTGAACCTCTTAACACCAACAGCTTATATTAAGTTT----- 89913
Qy 1120 TAGTAGGTGCTCAATAAATAATTTGATGAATGCACAGCTAGATATAAATTTCTTTTCT 1179
Db 89912 -----TGTGTTGTTTTAGAGACAGGCTCTCGCTCTCCACCCAGGCTTGAG 89867
Qy 1180 TTTTTTAAACAATCTTGACAACCTTTGCAAGATAAATAACAATCTTGCAATCTGCTTTTC 1239
Db 89866 TGGCTGCGCACAAATCATAGCTCACTAGGACTCAA-----GTGA 89828
Qy 1240 ACTTATCACTGTTATGACTTTTTCATATGTCCTCAAACTTTATGTTTACTGTTTTT 1299
Db 89827 TCCTCTGCTCTGAGTAACTGGGACTACAGGCTTGAGACACCATGCCCCAGAGATAG 89768
Qy 1300 CATTTGTTACTATTTTGTGCTGATATATGCTTAAATTTGCTTATATACATCTCTCTGCT 1359
Db 89767 CTCTCGATTGTTGCTTGGCTTTATACATTTCCACCTTGATGCTACTCAAAATGTTTT 89708
Qy 1360 CCATTTAGAGGCCAAATTTCAAACTGATGAAGCTATGAACCTCTCCCCAGAGAA 1419
Db 89707 ATTAGTCATA-----AATCTTTCTAATCACAATTTGCTCACTACAATTTCTATCA 89653
Qy 1420 ATACACACACACACACACTCACACAGTTTTTTTTTAAATGTTGCAACTAAGACAAG 1479
Db 89652 CAATTTACAGACTTTCAAAAATTTGATTTGTTTCTACTGAAATAAAACGTAGAGGAAAG 89593
Qy 1480 AAACCTGCA---TTAGAGGATTTTGTTCATATTAATTAATAAATCACTCAGTTGGGCACA 1536
Db 89592 AACACAGTAAGCTAAATGATGGAATTCATCCCTTAAATAAACAATTTACGCGGCGACG 89533
Qy 1537 GTGACTCAAGCTCTTAACACAGTACTTTTGAAGTCCAAGGTGGTGGATCACTTTGAGT 1596
Db 89532 TTGGCTCAGGCTGTATCTCAGACACTTTTGGAGGCCAGGTAGTGGATCAC--GAGGT 89475
Qy 1597 GAGAAGTTGAGACACAGCCTGGTCAATATGTTGAAACCTCTATCTCTACTAAAAATACAAA 1656
Db 89474 CAGGGCTCGAGACACAGCCTGACCAACATGTTGTAACCCCTCTCTACTAAAAATACAAA 89415
Qy 1657 AATTAGCTGGGTGAGTATGATGATGCTGTAGTCCAGCTACTCGGAGGCTGAGGCAAG 1716
Db 89414 AATTAGCTGGGTGAGTATGATGATGCTGTAGTCCAGCTACTCGGAGGCTGAGGCAAG 89355
Qy 1717 AGAATTCCTTGAACCTGGGAGGAGAGGTTGAGTGGAGCGGAGATCCACCACTGCACTC 1776
Db 89354 AGAATTCCTTGAACCTGGGAGGAGGAGGAGTGGAGTGGAGCGGAGATCCAGCTATCT 89295
Qy 1777 CAGCTGGGCGACACAGCGAGCTCTATCTCAAAAAATAAATAAATAAATA 1829
|||||

Db 89294 CAGCTGGGCGACAGCCAGCTCCGCTCTCAAAAAACAAAAAACA 89242
RESULT 7
AC009244
LOCUS 190162 bp DNA linear PRI 26-MAY-2002
DEFINITION Homo sapiens chromosome 7 clone RP11-305M3, complete sequence.
AC009244
VERSION AC009244.25 GI:21217399
KEYWORDS HTG.
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini;
Hominidae; Homo.
REFERENCE 1 (bases 1 to 190162)
AUTHORS Kaul,R.K., Olson,M.V., Zhou,Y., James,R.A., Rouse,G., Wu,Z.,
Saenphimmachak,C., Phelps,K.A., Buckley,D., Kibukawa,M., Raymond,C.
and Haugen,E.D.
Direct Submission
Unpublished
REFERENCE 2 (bases 1 to 190162)
AUTHORS Bubb,K.L., Desmarais,C.L., Ramsey,S.A. and Hubley,R.M.
Direct Submission
TITLE Submitted (07-AUG-1999) Human Genome Center, University of
JOURNAL Washington, Box 352145, Seattle, WA 98195, USA
REFERENCE 3 (bases 1 to 190162)
AUTHORS Kaul,R.K., Zhou,Y., James,R.A., Raymond,C., Haugen,E.D. and
Olson,M.V.
Direct Submission
TITLE Submitted (30-OCT-2000) Genome Center, University of Washington,
JOURNAL Box 352145, Seattle, WA 98195, USA
REFERENCE 4 (bases 1 to 190162)
AUTHORS Kaul,R.K., Olson,M.V., Zhou,Y., James,R.A., Rouse,G., Wu,Z.,
Saenphimmachak,C., Phelps,K.A., Buckley,D., Kibukawa,M., Raymond,C.
and Haugen,E.D.
Direct Submission
TITLE Submitted (26-MAY-2002) Genome Center, University of Washington,
JOURNAL Box 352145, Seattle, WA 98195, USA
COMMENT On May 26, 2002 this sequence version replaced gi:11038533.
----- Genome Center
Center: University of Washington
Center Code: UWGC
Web site: http://www.genome.washington.edu
Contact: uwgctg@u.washington.edu
----- Project Information
Center project name: chr-7
Center clone name: RP11-305M3 (djs195)
----- Summary Statistics
Assembly program: Phrap; version 0.990319
Consensus quality: 189969 bases at least Q40
Consensus quality: 190110 bases at least Q30
Consensus quality: 190161 bases at least Q20
Insert size: 190162; sum-of-contigs
Quality coverage: 9.7x in Q20 bases; sum-of-contigs

Overlapping Sequences:
5': RP11-382M23 (UMGC:djs734g) AC093149 1340-bp overlap
3': RP11-448A19 (UMGC:djs705) AC078846 8765-bp overlap

Sequence Quality Assessment:
This entry has been annotated with sequence quality estimates computed by the Phrap assembly program.
All manually edited bases have been reduced to quality zero.
Quality levels above 40 are expected to have less than 1 error in 10,000 bp.
Base-by-base quality values are not generally visible from the GenBank flat file format but are available as part of this entry's ASN.1 file.

This sequence was finished as follows unless otherwise noted:
all regions were either double-stranded or sequenced with an alternate chemistry or covered by high quality data (i.e., Phred

quality >= 30); an attempt was made to resolve all sequencing problems, such as compressions and repeats; all regions were covered by at least one plasmid subclone or more than one M13 subclone; and the assembly was confirmed by restriction digest.

Sequence Validation:

This sequence has been validated by Multiple Complete Digest fingerprinting. Comparison of the experimentally derived digest fragments with sequence-predicted fragments is given below. The electronically-digested sequence consists of both insert and vector, in order to accurately represent the entire circular BAC. Small fragments below a variable cutoff (approximately 400-800 bp) are not resolved in the fingerprint and hence do not appear in the table. There are no significant remaining discrepancies between the experimental and predicted values. Uniquely ordered fragments are separated by dashed lines.

SeqDerMap FngPrnt SeqDerMap FngPrnt SeqDerMap FngPrnt
 Nsil BglII EcoRI

21251	20929	12417	12174	8696	8794
12148	11918	2067	2041	6	<800
199	<800	5666	5826	1775	1773
3275	3221	254	<800	2000	1991
20379	20929	3818	3836	6313	6324
15555	15150	747	735	3256	3226
675	<800	9196	9237	325	<800
5493	5475	4154	4101	12989	12723
516	<800	5820	5826	5955	5951
5810	5831	118	<800	611	<800
9488	9435	2524	2359	1995	1991
2908	2935	8844	8768	11945	12009
779	840	332	<800	5577	5593
821	840	564	<800	4139	4124
315	<800	26	<800	955	940
3640	3671	4675	4644	2961	2940
428	<800	1801	1803	5264	5205
1144	1119	117	<800	1535	1534
4684	4717	6797	6799	3113	3383
20	<800	3146	3177	3760	3761
492	<800	2376	2359	3061	3110
10935	10879	210	<800	9605	9505
5921	6010	5135	5110	335	<800
529	<800	744	735	2190	2194
4354	4384	881	947	5117	5205
1311	1293	413	<800	13558	13434

10	<800	6389	6456	9367	9505
374	<800	296	<800	8080	8019
1497	1480	7160	7119	4276	4243
4079	4083	4846	4822	7830	8019
337	<800	2914	2924	231	<800
4483	4384	3803	3836	533	<800
1530	1480	1832	1803	668	<800
249	<800	674	<800	12314	12009
8831	8995	2515	2513	18561	18820
244	<800	2163	2137	11006	11013
1019	1000	466	<800	8756	8794
841	840	927	947	7	<800
8114	8086	5963	5826		
9184	9435	7440	7660		
7121	7263	5159	4822		
1122	1119	677	<800		
49	<800	1399	1373		
11928	11918	1565	1554		
1499	1480	2388	2513		
2585	2575	4826	4822		
699	<800	7661	7660		
		3624	3705		
		8693	8768		
		1749	1803		
		1642	1635		
		5295	5362		
		10628	10715		
		3227	3177		
		589	<800		

Query Match 23.1%; Score 462.6; DB 8; Length 190162;
 Best Local Similarity 61.3%; Pred. No. 2.3e-85;
 Matches 900; Conservative 1; Mismatches 520; Indels 46; Gaps 8;

QY 404 TTTTCTTTTGTGTTTGTAGACAGGGTCTTGTCTGTCTACCCAGGATGACACACT 463
 Db 31687 TTTTCTTTTAAAAAGTAGAAACGGAGTCTTGTGGGTACCCAGGCTGGGGTGCAGT 31746
 QY 464 GTTGCACCATAGTCTACCTGAGCTCAACCTCCTGAGCTCAAGGATCTGTGACCTCA 523
 Db 31747 AGTGCAGTCGTGTTTCACTGCATCTTCAATTCCTAGGAACAGGTGATCTCCACTTCA 31806
 QY 524 GCCTCCCAAGTAGCTGGGACTACGAGCGTGCACCACACGCTGGCTAATTAATAAAATT 583

Db 31807 GTCTTCTGAGTGGGATCTTTCAGGCATGTGCCCATGCTTAAATTTAAAAA 31866
QY 584 TTTTGTGAGACTGGGTCTTACTAGTTGGCCAGCTTGTCTTAAACTCTGGCTTCAA 643
Db 31867 TGTGTGAACACAGGGTCTTCTGTGTGTCTCAGGCTGTCTCAAAGCAGCCTCAA 31926
QY 644 GCATCTCTCTACCTTGGCATCCCAAGGTCTGGGATTAAGGGGTGAGCCACCATGTGC 703
Db 31927 GTGATCTCTCTGCTCATCTCCCAAGGTCTGGGATTAAGGGGTGAGCCACCATGTGC 31986
QY 704 GGTACTTATTTCTTACATTCATCTTCTCAATAGATGTAAGATCCACAGACAGGA 763
Db 31987 AGCCCATTTCTTTTCTGTTTTTTTTTTTAAATTTAATTTTATTTTATTTT 32046
QY 764 TTACTGCTTATTTCTTCTCTT-TCTTTTGTGAGACAGAGTCTCACTTCACTCACTCAACC 822
Db 32047 TATTTTATTTATTTATTTATTTTGTGAGATGAAGTCTCACTCTGTGCCCGCT 32106
QY 823 TCCGT-----TCAGCTCATGTCAACCTCTGCTTCCGGGTCAAGYATTC 868
Db 32107 GGAGTGCAGTGATCTCGGCTCACTGAAACCTCTGCTTCCAGGTTCAGGATTC 32166
QY 869 TCCTGCTAAGCTCTCAGTAGCTGGAATTACAGCGTGCACACCATGCTTGGCTTAAT 928
Db 32167 TCTTGCCTCAGCTCTCAGTAGCTGGGATTAAGGTCCTCCCATATACCCGGCTAA- 32225
QY 929 TTTTGTATTTTACAGAGATGGGTTTTTACCATGTTGCCAGGCTGTCTCAAACTCC 988
Db 32226 TTTTGTATTTTATAGAGACGGGTTTTTACCATGTTAGTACGCTGTCTCAAACTCC 32285
QY 989 TGACTCAAGTGAATCTGCTGCTCAGTCTCCAAAGTCTGGGAATTATAGCGGTGAGTC 1048
Db 32286 TGACTCAGATGATCTGCTGCTGCTGCCCTCCAAAGTCTGGGATTAAGGTCGAGTC 32345
QY 1049 ACTGTGCTGCGGATTAATGTTATTTT-----CTTATGCTATATCCCAAGATAGAG 1105
Db 32346 ACCGACCCGGTCCCAATTTTTTTTTTAAACAAATGCTTACTGACCTACGCTAAGAG 32405
QY 1106 CAGTGTCTGACATAGTAGGTCTCAATAATATTGATGAATGCACAGCTAGATATA 1165
Db 32406 CTGTAAGGCGAATGAGACATGCTTACCTCAGAACTTATATTTTGTGAGATAT 32465
QY 1166 AACTTTCTTTTCTTTTTTAAACAACTCTTGACAACTTTGCAGAAATAATACAACTCTG 1225
Db 32466 ACACATATATAAATGATGATACAGGACATATATAAATGCCATGACAGATTATA 32525
QY 1226 CATCTGCTTTTTCAC-TTATCACCTGTTATGACTTTTTCATATTTGCTCAAACTTTA 1284
Db 32526 AATCATATTCACACACAAAGAGGCTTAGGAAAGTTCCAGATTCCCACTTTGGGAAG 32585
QY 1285 TTGTTACTGTTTTTCTTATTTTACTATTTTGTGCTCAATTAATATGCTTAAATTTGCTT 1344
Db 32586 CTGAGGCGAGGTGATCCTGAGGTGAGGAATTC--AGACCAGGTGGCCCAACATGGT 32643
QY 1345 ATACATCTCTGCTTCCACTTTGAAGGCCAAATTTACAAATCTGATGAAGCTATGAAC 1404
Db 32644 AAACCTGCTTTTACTAGAGTACAAAATTTAGTGTGAGTGGTGGTGCAGGCTGTAAAT 32703
QY 1405 CCTCTCCCGAGAGAAATACACACACACACACACTACACAGTTTATTTTAAATGTT 1464
Db 32704 CCTAGCTACTCATGAGGCTGAGGCGAGGAATCGCTTGAACCCAGGAGGCGGCTGCA 32763
QY 1465 TGCACCTAAGACAGAACTGCTATAGAGATGTTG----- 1502
Db 32764 GTGAGCGAGATCGCGCACCTGCACTTAGCTGGGTGACAGGTGAGACTTCAATCTCAA 32823
QY 1503 TTTATATTAATTAATAATACTCAGTTGGGCACAGTGAAGTCAAGCTGTGAACACAGTAC 1562
Db 32824 AAAAAAAAAAAAAAAAAAGAGCGCGCGGTGCTCAGCTTGTATTTCCAGCAC 32883
QY 1563 TTTGGAAGTCAAGGTGGTGGATCACTTGAAGTGAAGTTTCAGACACCAAGCTGTGTTCAA 1622

Db 32884 TTTGGGAGCCTTGAGTGGCGGATCA--TGAGGTGAGAAGTTTCAGACACAGCTGGCCAA 32941
QY 1623 TATGTTGAAACCTTATCTTACTTAAATAACAAATACTAGCTGGTGTAGTGTGATGTCATGC 1682
Db 32942 TATGTTGAAACCCGCTCTTACTTAAAGTACAAATACTAGCTGGTGTGCGCCAGCAC 33001
QY 1683 CTGTAGTCCCAAGCTTACTGGGAGGCTGAGGCAAGAGAATCTTTGAACCTGGGAGGCAGA 1742
Db 33002 CTGTAGTCCCAAGCTTACTCAGGAGCTTAAGGCAGGAGAATCGCTTGAACCCAGGAGGCAGA 33061
QY 1743 GGTGTGAGTGGCCGAGATCCCACTGCTGCTCAGCTGGGCGACACAGCGAGACTCT 1802
Db 33062 GGTGTGAGTGGCCGAGATTTATGCTGCTGCTCAGCTGGTGTGACAGAGCAAGACTCC 33121
QY 1803 ATCTCAAAAAATAATAATAATA 1829
Db 33122 GTCTCAAAAAATAATAATAATA 33148

RESULT 8
AP000589
LOCUS Homo sapiens chromosome 11 clone CMB9-105NS map 11q13, WORKING
DEFINITION DRAFT SEQUENCE, 19 unordered pieces.
ACCESSION AP000589
VERSION AP000589.3 GI:8118795
KEYWORDS HTG; HTGS PHASE1; HTGS_DRAFT.
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini;
Homnidae; Homo.
REFERENCE 1 (bases 1 to 123192)
AUTHORS Hattori,M., Ishii,K., Toyoda,A., Taylor,T.D., Hong-Seog,P.,
Fujiyama,A., Yada,T., Totoki,Y., Watanabe,H. and Sakaki,Y.
TITLE Homo sapiens 123,192 genomic DNA of 11q13
JOURNAL Published Only in DataBase (1999)
REFERENCE 2 (bases 1 to 123192)
AUTHORS Hattori,M., Ishii,K., Toyoda,A., Taylor,T.D., Hong-Seog,P.,
Fujiyama,A., Yada,T., Totoki,Y., Watanabe,H. and Sakaki,Y.
TITLE Direct Submission
JOURNAL Submitted (12-OCT-1999) Masahira Hattori, The Institute of Physical
and Chemical Research (RIKEN), Genomic Sciences Center (GSC);
Kitasato Univ., 1-15-1 Kitasato, Sagamihara, Kanagawa 228-8555,
Japan (E-mail:hattori@gsc.riken.go.jp)
URL:http://hgp.gsc.riken.go.jp/, Tel:81-42-778-9923,
Fax:81-42-778-9924)
COMMENT On May 31, 2000 this sequence version replaced gi:6997479.
----- Genome Center
Center: RIKEN Genomic Sciences Center(GSC)
Center code: RIKEN
Web site: http://hgp.gsc.riken.go.jp/
Contact: hattori@gsc.riken.go.jp
----- Project Information
Center project name: HumDraft11
Center clone name: CMB9-105NS
----- Summary Statistics
Sequencing vector: PCR products; 100% of reads
Chemistry: Dye-terminator ET-amersham; 100% of reads
Assembly program: Phrap; version 0.990329
Consensus quality: 115116 bases at least Q40
Consensus quality: 118340 bases at least Q30
Consensus quality: 120191 bases at least Q20
Insert size: 121392; sum-of-coverage
Quality coverage: 5.08x in Q20 bases; sum-of-coverage

NOTE: This is a 'working draft' sequence. It currently consists of 19 contigs. The true order of the pieces is not known and their order in this sequence record is arbitrary. Gaps between the contigs are represented as runs N, but the exact sizes of the gaps are unknown. This record will be updated with the finished sequence as soon as it is available and the accession number will be preserved


```
QY 817 TCAACCTCGTTGCTCAGTCTCACTGCAACCTCTGCTCCCGGTTCAAGYGAATCTCTGCT 876
Db 63537 ATGGCGCAATCTCGGCTCACTGCAAACTCTGCTTCGAGGTTCAAGGATCTCTGCTGCT 63596
QY 877 AAGCTCTCTGAGTCTGCAATTAACAAGCTGCACCAACCATCTGCTGCTTAATTTT--G 934
Db 63597 GAGCTTCAGAGTAGTGGGATTAACAGGTGCGCGCCACACCATCTGCTGCTACTTTTGT 63656
QY 935 TAATTTTACAGAGATGGGTTTATACCATGTTGCCAGGCTGCTCAAACTCTCTGACT 994
Db 63657 ATTTTATAGACAGAGGTTTACCAATGTTGTCAGGCTGCTCTGAACTCTCTGACT 63716
QY 995 CAAGTATCTGCTGCTCAGTCTCCCAAGTCTGGAATTAAGGGTGAAGTCACTGTG 1054
Db 63717 CAGGTGATCGCTGCTCAGCTTCCAAAGTCTGGGATTAACAGGCATGAGCCACCGCG 63776
QY 1055 CCGGCGGATTAATCTCTATTTTCTTTATGCTATATCCCGAGTCTAGACAGTGTCTG 1114
Db 63777 CCGGATGCCCTTGAATCTTTGAAACAAATGTATATAAATTAATAATGCTAAACGT-TA 63835
QY 1115 ACATATAGTAGTGTCTCAATAATAATTAATGATGAATGCACAGCTAGATATAAACTTCTT 1174
Db 63836 TCATTTCTGTGATCATAGGCTCTGACACATTTTCTCTAGATATGTTGTCATATG 63895
QY 1175 TTTCTTTTAAACAATCTTGACAACCTTTGCGAGAATAATACAACTTTCGATCTGCT 1234
Db 63896 ATTAATATTCACATAAAGCAAAATAATATGGGTACTACAAATTTTGATGCTATTTAT 63955
QY 1235 TTTTCA-----CTTATCACCTGTTATGACTTTTTCATATGCTCTCAACCTTAT 1285
Db 63956 ATATCAATATAAGATGATTAGCCAGAGAAACATAATGTTATTTGGAAAGGTATCTGAA 64015
QY 1286 TGTTACTGTTTTTTCATTTGTTACTATTTTAGTCACTGAATAATATGCTTAATTTGCTTA 1345
Db 64016 TGGGTGTTTAGTTTTTCCCTCTCACTTAGTTAAATACACAGATGGAATATTTTGT 64075
QY 1346 TACATCTCTGCTCCACTTTAGAGGCCAAATTTTACAAATCTGATGAAGCTATGAACC 1405
Db 64076 TTCTAGAAACAAGACAGGAATAGCATCTGTTTTCATTTTATAGATATGACCACTGAGT 64135
QY 1406 CTCCTCCAGAGAAATACACACACACACACTCAACAGT-TTTTTTTTAAAGTT 1464
Db 64136 CCAATGAGCTAGAAGGAATTTGTTAGAGTGAATTTTCCCACTCTGCTCAAGTTGGAGGC 64195
QY 1465 TGCAACTAAGACAGAAACCTGCAATAGAGGATGTTTCTCATATTAATTAATAATAACT 1524
Db 64196 CAAAAATGTCAAATGTAATCTAGAGTCAAAATATTTTACTTTTTTAAATGATGTGA 64255
QY 1525 CAGTTGGGCA CAGTGACTCAAGCCCTGTAACCAACACAGTACTTTGGGAAGTCCAAAGTGGTGG 1584
Db 64256 AGGCCAGGTGGGTGGCTCAGTATGTAATCCAAGCACTTTGGGAAGCCGAGGCGAGTGG 64315
QY 1585 ATCACTTGAGTGAAGTTTCAGACAGCCTGCTCAATATGTTGAACCCCTATCTCTAC 1644
Db 64316 ATTACTTGAGTCAAGGAGTTTCAGACAGCCTGCGCAACATGTTGGAACCACTGTCTCTAC 64375
QY 1645 TAAAAATAAAAAATTAGCTGGGTGCTAGTATGATGATGCTCTAGTCCAGCTACTCTCGGA 1704
Db 64376 TAAAAATACAAAATTTAGCTGGGTGGTGTGATGCTGCTGTAAATCCAGCTACTCTGGA 64435
QY 1705 GGCTGAGCAAGAGAAATTTGTAACCTGGGAGGAGAGGTTGAGTGAGCCGAGATCCC 1764
Db 64436 GGCTGAGGCAAGAAATCGCTTGAACCGGGAGGAGGTTGAGTGAGCCGAGATGTT 64495
QY 1765 ACCTGAGCTCCAGCTCGGCGGACACAGCGAGACTCTATCTCAAAAAATAAATAAATA 1824
Db 64496 GTCACTGCACTCCAGCTCGGCGGCAACAGACCAAGACTTCTCTCAAAAAATAAATAAATA 64555
QY 1825 AAATAAAGGATCGGAGAGAAACAACTAATAAGA 1859
Db 64556 AAATAAATGAATAATAAATAATGATGTTAAAGATA 64590
```

```
RESULT 9
AC022408
LOCUS Homo sapiens chromosome 15 clone RP11-118J16 map 15q15, complete
DEFINITION AC022408
AC022408
AC022408.6 GI:15217198
HTG.
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini;
Hominidae; Homo.
REFERENCE 1 (bases 1 to 150266)
AUTHORS Rowen,L., Madan,A., Qin,S., Baradarani,L., Birditt,B., Bloom,S.,
Burke,J., Dors,M., Fleetwood,P., Kaur,A., Madan,A., Nesbitt,R.,
Pate,D. and Hood,L.
TITLE Sequencing of human chromosome 15 D15S146-D15S117 region
JOURNAL Unpublished
AUTHORS 2 (bases 1 to 150266)
Rowen,L., Madan,A., Qin,S., Abbasi,N., Baradarani,L., Birditt,B.,
Bloom,S., Dors,M., Dickhoff,R., Fleetwood,P., Harrison,G.,
James,R., Kaur,A., Madan,A., Owen,M.P., Ratcliffe,A., Shaffer,T.
and Hood,L.
TITLE Direct Submission
JOURNAL Submitted (03-FEB-2000) Multimegabase Sequencing Center, University
of Washington, PO BOX 357730, Seattle, WA 98195, USA
AUTHORS 3 (bases 1 to 150266)
Rowen,L., Madan,A., Qin,S., Baradarani,L., Birditt,B., Bloom,S.,
Burke,J., Dors,M., Fleetwood,P., Kaur,A., Madan,A., Nesbitt,R.,
Pate,D. and Hood,L.
TITLE Direct Submission
JOURNAL Submitted (21-AUG-2001) Multimegabase Sequencing Center, Institute
for Systems Biology, 4225 Roosevelt Way NE, Suite 200, Seattle, WA
98105, USA
COMMENT On Aug 21, 2001 this sequence version replaced gi:11225361.
----- Genome Center
Center: Multimegabase Sequencing Center
Web site: http://chroma.mbt.washington.edu/meg_www
Contact: leerowen@systemsbiology.org
----- Summary Statistics
Sequencing vector: pUC18; L08752
Chemistry: Dye-terminator Big Dye; 90% of reads
Chemistry: Dye-primer Big Dye; 10% of reads
Assembly program: Phrap; version 0.990399
Note: data from AC021753 [Drafting center UWMSC] and AC012652
[Drafting center UWMSC] and 58017 [Drafting center WUGSC] were
added for finishing.
FEATURES
source 1..150266
/organism="Homo sapiens"
/mol_type="genomic DNA"
/db_xref="taxon:9606"
/chromosome="15"
/map="15q15"
/clone="RP11-118J16"
/clone_lib="RPCI human BAC library 11"
/notes="This clone overlaps RP11-46M12 AC012652 and
RP11-129112 AC021753 and RP11-58017 AC068981.Data from
overlapping BACs were combined and the consensus sequence
determined from RP11-118J16."
misc_feature 1..150266
/notes="overlap with RP11-58017 AC068981."
misc_feature 1..100397
/notes="overlap with RP11-129112 AC021753."
unsure 49355..49359
/notes="low quality data."
misc_feature 100628..150266
/notes="overlap with RP11-46M12 AC012652."
ORIGIN
Query Match 23.0%; Score 460.4; DB 8; Length 150266;
```

Db	53106	CAATTTACAGACTTTCAAAAAATTGATTTGGTTCTTACTGAAATAAAACGTAGAGGAAAAG	53165
Qy	1480	AAACCTCGCA---TTAGAGGAGTGTGTTTCATATTAATTTAAAAAATAACTCAGTTGGGCACA	1536
Db	53166	AACACAGTAGCTTAAATGATGGAATTCATCCCTTAAAAACAAAAATTTCAGCCGGCAGC	53225
Qy	1537	GTGACTCAAGCCCTGTTAAACACACAGTACTTTTGGAAAGTCCAAAGGTGGGTGGATCACTTTGAGGT	1596
Db	53226	GTGGCTCAGCCCTGTAATCTCAGCACTTTTGGGAGGCCAAGTAGGTGGATCAC--GAGGT	53283
Qy	1597	GAGNAGTTTCGAGACAGCCCTGGTCAATATGTTGAAACCCCTACTCTACTTAAAAATAACAAA	1656
Db	53284	CAGGGGCTCAGAGACCAGCCCTGACCAACATGTTGTAACCCCGCTCTCTATAAAATAACAA	53343
Qy	1657	AAATTAGCTGGGTCTAGTGAATGATGCTGCTCTAGTCCAGACTACTCGGAGGCGCTGAGGCAG	1716
Db	53344	AAATTAGCTGGGTCTGTTGGCGGCGCCTGTGTAATACCAGCTACTCAGGAGGCTGAGGCAG	53403
Qy	1717	AGAATTCGTTGAACCTTGGGAGGCAGAGGTTGTCNAGTCAGCCGAGATCCCAACCACTGCATC	1776
Db	53404	AGAATTCGTTTGAACCCCGGAGCGCGAAGCTGTCAGTCAGCCGAGATCACACCACTATATCTC	53463
Qy	1777	CAGCCTGGGCGACACAGCGAGACTCTATCTCAAAAAATAAATAAATAAATAA	1830
Db	53464	CAGCCTGGGCGACAGAGCCAGACTCCGCTCTCAAAAACACAAAAAATAAATAA	53517

```

RESULT 10
AC021753      167996 bp      DNA      linear      PRI 29-OCT-2001
LOCUS        Homo sapiens chromosome 15 clone RP11-129I12 map 15q15, complete
DEFINITION
ACCESSION    AC021753
KEYWORDS
SOURCE       HTG.
ORGANISM     Homo sapiens (human)
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini;
Hominidae; Homo.
1 (bases 1 to 167996)
Rowen,L., Madan,A., Qin,S., Baradarani,L., Birditt,B., Bloom,S.,
Burke,J., Dors,M., Fleetwood,P., Kaur,A., Madan,A., Nesbitt,R.,
Pate,D. and Hood,L.
Sequencing of human chromosome 15 D15S146-D15S117 region
Unpublished
2 (bases 1 to 167996)
Rowen,L., Madan,A., Qin,S., Abbasi,N., Baradarani,L., Birditt,B.,
Bloom,S., Dors,M., Dickhoff,R., Fleetwood,P., Harrison,G.,
James,R., Kaur,A., Madan,A., Owen,M.P., Ratcliffe,A., Shaffer,T.
and Hood,L.
Direct Submission
Submitted (20-JAN-2000) Multimegabase Sequencing Center, University
of Washington, PO BOX 357730, Seattle, WA 98195, USA
3 (bases 1 to 167996)
Rowen,L., Madan,A., Qin,S., Baradarani,L., Birditt,B., Bloom,S.,
Burke,J., Dors,M., Fleetwood,P., Kaur,A., Madan,A., Nesbitt,R.,
Pate,D. and Hood,L.
Direct Submission
Submitted (29-OCT-2001) Multimegabase Sequencing Center, Institute
for Systems Biology, 4225 Roosevelt Way NE, Suite 200, Seattle, WA
98105, USA
On Oct 29, 2001 this sequence version replaced gi:11136698.

----- Genome Center
Center: Multimegabase Sequencing Center
Contact: leerowen@systemsbiology.org
Web site: http://chroma.mbt.washington.edu/msg_www
----- Summary Statistics
Sequencing vector: pUC18; L08752
Chemistry: Dye-terminator Big Dye; 90% of reads
Assembly: Dye-primer Big Dye; 10% of reads
Assembly program: Phrap; version 0.990399

```


Note: data from AC022408 [Drafting center UWMSC] and AC020661 were added for finishing.

FEATURES
source
1. .167996
/organism="Homo sapiens"
/mol_type="genomic DNA"
/db_xref="taxon:9606"
/chromosome="15"
/map="15q15"
/clone="RP11-129112"
/clone_lib="RPC1 Human BAC library 11"
/note="This clone overlaps RP11-118J16 AC022408 and RP11-540U11 AC020661. Data from overlapping BACs were combined and the consensus sequence determined from RP11-129112 to the extent possible."
misc_feature
1. .63402
/note="overlap with RP11-540U11 AC020661."
misc_feature
67596..167996
/note="overlap with RP11-118J16 AC022408."

ORIGIN

Query Match 23.0%; Score 459.4; DB 8; Length 167996;
Best Local Similarity 62.0%; Pred. No. 1.1e-84;
Matches 888; Conservative 1; Mismatches 487; Indels 57; Gaps 8;
Qy 400 TGGGTTTTTTTGTGTTTGTAGACAGGGTCTTCTCTGTCAACCCAGGCGATGACGA 459
Db 119755 TGTGTAATTTTGTGTTGTTGTTGAGCCAGCTCGGCTCTGTGCGCCAGGCTGGAGTG 119814
Qy 460 CAGTGGTGCACCATAGTCTACTGAGGCTCAACTCTCTGAGCTCAAGGATCTGCTGAC 519
Db 119815 CAGTGGTGCATCTCGGCTCACTGCAATCTCTGCTCTCGGGCTCAAGCCATCTCTCAC 119874
Qy 520 CTCAGCCTCCCAAGTAGTGGGACTACGAGGTGCACACCAACCGCTGGCTAAATTA 579
Db 119875 CTCAGCATCCCAAGTAGTGGGACTACGAGGTGCATCACATGCTTAGCTAATTTTGT 119934
Qy 580 AATTTTTTTTGTAGACAGGCTTTACTACGTTGGCCAGGCTTGTCTTAAACTCTCGCT 639
Db 119935 A--TTTTCTGTAGACAGGCTTTTGGCCTGTGCCCAGTCTGATTTTCAAACTCTGAGC 119992
Qy 640 TCAGCATCTCTCTACTGTCATCCCAAGTCTGGGATTAAGGGGTGAGCCACAT 699
Db 119993 TCAACAGCATCGCCT-GCTCAGCCTCCCAAGTCCGGGATTAAGGGTGTGAGCCACAA 120051
Qy 700 GTGCGGCTACTTATTTCTTACATTCATCTTCCATAGATTAAGATCCACAGACA 759
Db 120052 GCCTAACCTATTAATGTGAATTTGTATGTTGGCATATAAAATCCCTAGCCAGTTTGT 120111
Qy 760 GGGATTACTGCTATTTCTTCTTCTTTTGTGAGACAGAGTCTCACTTCATCACCTCA 819
Db 120112 TTTGTT-TTGTGTTTGTGAGACAGATTTTGTCTTGTGCACTCCAGTGAGTGAGTG 120170
Qy 820 ACTCCCGTCACTGCACTGCAACTCTGCTCCCGGTTCAAGYGATTTCTCTGCTCAAG 879
Db 120171 GCTCAATCTCGGCTCACTGCAACTCTCGGCTCCCGGTTCAAGGATTTCTCTGCTCAG 120230
Qy 880 CCTCTCAGTAGTGGGATTAAGGCTGACACCATGCTTGGCTAATTTTGTATTT 939
Db 120231 CCTCTCAGTAGTGGGATTAAGGCTGACACCATGCTTGGCTAATTTTGTATTT 120290
Qy 940 TTAGCAGAGATGGGTTTACATGTTTGGCCAGGCTGCTCAAACTCTGACCTCAAGT 999
Db 120291 TTGTAGAGTCAGATTTTAACTGTTGGCCAGGCTGTTTGGCACTCTGACCTCAGT 120350
Qy 1000 GATCTGCTGCTCAGTCTCCCAAGTCTGGAATTAAGCGGTGAGTCACTGTGCTGG 1059
Db 120351 GATCCACCTGCTCAGTCTCCCAAGTCTGGAATTAAGCGGTGAGTCACTGTGCTGG 120410
Qy 1060 CCGATTACTCTATTTTCTTATTTGCTATATCCAGATCTAGACAGTGTCTGACAT 1119
Db 120411 CCAACAAACTTTCATTAAGTGAACCTCTTAAACACACAGCTTATAATTAGTT----- 120462

Qy 1120 TAGTAGGTGCTCAATAAATAATGATGAATGCAAGCTAGATATAAATTTCTTTTCT 1179
Db 120463 -----TGTTTTGTTTTTAGAGACAGGGTCTCGCTCTCCACCCAGGCTTGAG 120508
Qy 1180 TTTTTPAAAACAATCTTGACAACCTTTGCGAGAATAAATAAATCTTGCAATCTGCTTTTC 1239
Db 120509 TGGCTGGCAACAATCATAGCTCACTAGGGACTCAA-----GTGA 120547
Qy 1240 ACTTATCACCTGTTATGACTTTTTCATATGTTGCTCAAACTTTATTTGTTACTGTTTTT 1299
Db 120548 TCCTCTGCTCTCTGAGTAAGTGGGACTCAGGCTGAGACACCATGCCCGCAGAAATAG 120607
Qy 1300 CATTTGTTACTATTTTACTGCTCAATGAATATGGCTTAATTTGCTTATACATCTCTCTGCT 1359
Db 120608 CTCTCGAATTTGTTGCTTTGCTTAATCAATTTCCACCTTGATGCTACTCAAAATGTTTT 120667
Qy 1360 CCATTTAGAAAGGCAATTTTACAATCTGATGAAGCTATGAACCTCTCCCCAGAGAA 1419
Db 120668 ATTAGTCATA-----AATCTTTCTTAATCAATTTGCTCACTCAATTTATTTCTATCA 120722
Qy 1420 ATACACACACACACACACTCTCACACAGCTTTTTTTTTTAAATGTTTGCACCTAAGACAAG 1479
Db 120723 CAATTTACAGACTTTCAAAATTTGATTTGTTGTTCTACTGAATAAACAAGCTAGAGGAAAG 120782
Qy 1480 AAACCTGCA---TTAGAGGATGTTTGTTCATATTAATTAATAAATAAATAAATAAATAA 1536
Db 120783 AACACAGTAAGCTAAATGATGAATTTCAATCCCTTAATAAACAATAAATAAATAAATAA 120842
Qy 1537 GTGACTCAAGCTGTAAACACAGTACTTTTGAAGTCCAGGTTGGTGGATCACTTTGAGGT 1596
Db 120843 GTGGCTCACGCTGTAATCTCAGACCTTTGGGAGGCAAGGTAGTGGATCAC--GAGGT 120900
Qy 1597 GAGAAAGTTCCAGACACAGCTGCTCAATATGCTGAACCTTCTCTACTAAAAATACAAA 1656
Db 120901 CAGGGCTCAGACACAGCTGACCAATGTTGAAACCTGCTCTACTAAAAATACAAA 120960
Qy 1657 AATTAGCTGGGTAGTGTATGCTGTAGTCTCCAGCTACTCTGGAGGCTCAGGCAAG 1716
Db 120961 AATTAGCTGGGTGTGGTGGCGGCGCTGTAAATACAGCTACTCAGGAGGCTCAGGCAAG 121020
Qy 1717 AGAATTTGTTGAACCTGGGAGGAGGAGGTTTGAAGTGGAGCGGAGATCCACACTGCACT 1776
Db 121021 AGAATTTGTTGAACCTGGGAGGAGGAGTGTGAGTGGAGCGAGATCAGGCACTATATC 121080
Qy 1777 CAGCCTGGGCGACACAGGAGCTCTATCTCAAAAAATAAATAAATAAATA 1829
Db 121081 CAGCCTGGGCGACAGAGGAGGAGTCTCGTCTCAAAAAATAAATAAATAAATAA 121133

RESULT 11

AC003689 137693 bp DNA linear PRI 21-NOV-1998
Homo sapiens Chromosome 11q12.2 PAC clone pdj1081b4 containing human mRNA for T-cell glycoprotein CD6, complete sequence.
AC003689
AC003689.1 GI:3900834
HTG.
Homo sapiens (human)
Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini; Hominidae; Homo.
1 (bases 1 to 137693)
Evans, G.A., Athanasiou, M., Aguayo, P., Armstrong, D., Basit, M., Buettner, J., Bumeister, R., Card, P., deSallabot, F., Dunn, J., English, C., Ethridge, S., Garner, H.R., Gee, V., Gordon, M., Gotway, G., Grant, O., McFarland, J., Joslin, J., Lewis, E., Loo, H., Loo, K.N., Major, T., McFarland, J., Newton, J., Osborne-Lawrence, S., Schageman, J., Schultz, R.A., Stimson, S., Syed, M. and Ward, T.
HTGS Submission
Unpublished
TITLE
JOURNAL
REFERENCE
2 (bases 1 to 137693)
Evans, G.A., Athanasiou, M., Basit, M., Bradbury, P., Brignac, S.,

repeat_region	complement (66479. .66870)	QY	935	TATTTTGTAGCAGAGATGGGGTTTACCATGTTGGCCAGGCTGGTCTCAAACTCTCTGACCT	994
repeat_region	/rpt_family="Alu"	Db	77109	ATTTTTTTAGTAGACACAGGGTTTCCCATGTTGGTCCAGGCTGGTCTCGAACTCTCTGACCT	77168
repeat_region	67594. .67887	QY	995	CAAGTGATCTGCCTGCCTCAGTCTCCCAAAGTGTGGAATTTATAGGCGGTAGTCACTGTG	1054
repeat_region	/rpt_family="Alu"	Db	77169	CAGGTGATCCGCTGCCTCAGCTTCCAAAGTGTGGGATTTACAGGCATGAGCCACCCGG	77228
repeat_region	68119. .68182	QY	1055	CCTGCGGATTAATGCTATTTTCTTATTTCTTATTCCTATATCCCCAGATCTAGACAGTGTCTG	1114
repeat_region	/rpt_family="MER42"	Db	77229	CCCGATGCCCTTTCATCTTTTGAACAATAATGATATAAATAAATGCTAAAGCT-TA	77287
repeat_region	complement (69234. .69519)	QY	1115	ACATATAGTAGGTCTCAATAAATAATTTGATGAATGCACAGCCCTAGATATAAACTTTCTT	1174
repeat_region	/rpt_family="Alu"	Db	77288	TCATTTCTGTGATCATAGGCTCTGACAACTTTTCTCTCTAGATTTATGTTGTCTATTG	77347
repeat_region	complement (69560. .69821)	QY	1175	TTTCTTTTAAAAACAATCTTGACAACTTTTGAGAAATAATAAATCAATCTTGCATCTGCT	1234
repeat_region	/rpt_family="Alu"	Db	77348	ATTATTTTACACATAAAGCAAAATAATATGGGTACTACAAATTTTGTATCTATTTAT	77407
repeat_region	complement (70420. .70676)	QY	1235	TTTTCA-----CTTATCACCTTGTATGACTTTTTCATATTTGCTCTCAAACTTTAT	1285
repeat_region	/rpt_family="MER42"	Db	77408	ATAATCAATATAAGGATGATAGCCAGAGAAACTAAATGTTTATGAAAGGTATCTGAA	77467
repeat_region	71247. .71537	QY	1286	TGTTACTGTTTTTTTCTTACTTATTTTAGTCTCACTGAATAATATGCTTAATTTGCTTA	1345
repeat_region	/rpt_family="Alu"	Db	77468	TGGGTGTTTAGGTTTTTCCCTCTCCTTCTAGTTAAATACCCAGAGATGATATTTTGT	77527
repeat_region	complement (76584. .76879)	QY	1346	TACATCTCTCTGCTCCACTTTTAGAAGGCCAAATTTTAAATCTCTGATGAAGCTATGAAC	1405
repeat_region	/rpt_family="Alu"	Db	77528	TTCTAGAAACAAGACAGGAATTAGCATCTCTGTTTTTCAATTTTATAGATATGACCACTG	77587
repeat_region	77719. .77994	QY	1406	CTCTCTCCAGAGAAATACACACACACACACTCACAACACT-TTTTTTTTAAATGTT	1464
repeat_region	/rpt_family="Alu"	Db	77588	CCATGGAGCTAGAAAGGAATTTTGTAGAGTGAATTTTCCCACTCTGCTCCAGTTGGAGGC	77647
repeat_region	22.9%; Score 457.4; DB 8; Length 137693;	QY	1465	TGCAACTAAGACAGAAACCTGCAATTAGAGGATGTTTGTTCATATTAATTAATAATAACT	1524
repeat_region	Best Local Similarity 59.7%; Pred. No. 2.8e-84;	Db	77648	CANAATGTCAATATGTAATCTAGAGTCAAAATTTTACTTTTTTAAATGATGTTGA	77707
repeat_region	Matches 880; Conservative 1; Mismatches 577; Indels 17; Gaps 6;	QY	1525	CAGTTGGGCACAGTACTCAAGCTGTGAACACAGTACTTTTGGAAAGTCCAAAGTGGGTGG	1584
repeat_region	398 TTTGGTTTTTTTTTTTGTGTTTGTAGACAGAGGTCTTGCTC-TGTCAACCAGGATGA	Db	77708	AGGCCAGTGGGTGGTCTCAGTATGTAATCCAAAGCACTTTTGGAAAGCCGAGGAGGTGG	77767
repeat_region	76572 TTCTTCTCTCTTTTTTTTTTTTGTAGACAGAGTTCTCTCTGTTGCCAGGCTGGA	QY	1585	ATCACTTGAAGTGAAGATTTGAGACAGCCTGTCAATATGTTGAACCTTATCTCTAC	1644
repeat_region	457 GCACAGTGTGAACCATAGTCACTGACGCTCAACCTCTGAGCTCAAGGGATCTGCT	Db	77768	ATTACTTGAGTCAAGGATTCGAGACAGCCTGGCCAAATGGTGAACCACTGCTCTCTAC	77827
repeat_region	76632 GTGCAATGGCAAACTCTCGGTTTCAACCAACCTCCGCTCCCGGTTTCAAGGGTTCTCC	QY	1645	TAAAAATACAAAATTAGCTGGGTGATGATGCTGCTGATGCTCCAGCTACTCGGGA	1704
repeat_region	517 GACCTCAGCTCCCAAGTAGCTGGGACTAGGAGCTGACCAACAGCCTGGCTAATTA	Db	77828	TAAAAATACAAAATTAGCTGGGCTGTGATGGTGCCTGTAATCCAGCTACTTGGGA	77887
repeat_region	76692 TGCCTCAGCTCCCGAGTAGCTGAGATTACAGGCATGCGCACCAAGCCGAGCTGATTT	QY	1705	GGCTGAGGCAAGAGAATTTGTAACCTGGAGGCGAGAGGTGTCAGTGAGCCGAGATCCC	1764
repeat_region	577 AAAAAATTTTTTGTAGAGACTGGGTCTTACTACGTTGGCCAGGCTTGTCTTAAACTCCTG	Db	77888	GGCTGAGGCAAGAGAATCGCTTGAACCCGCGGAGCGAGGTGTCAGTGAGCCGAGATGT	77947
repeat_region	76752 GTA---TTATTAGTAGACAGCGGTTTCTCCATGTTGTCAGGCTGCTCTCAGCTCTG	QY	1765	ACCACTGCATCTCAGCTGGCGCACACAGCACTCTATCTCAAAAATAAATAAATA	1824
repeat_region	637 GCTTCAAGCAATCTCTTACCTTGGCATCCAAAGTGTGGATTTACAGGGGTAGCCAC	Db	77948	GTCACTGCATCTCAGCTGGGCAACAGCAAGATTTCTGCTCAAAATTAATAAATA	78007
repeat_region	76809 ACCTCAGGTGATCCACCCACCTCGGCTCCCAAAGTGTGGATTTATAGTGTGAGCCAT	QY	1825	AAATAAAGGATCGAGAGAAACAAACTAATAAGA	1859
repeat_region	697 CATGTGGGCTACTTATTTCTTCAATTCATCTTCCATAGATGTAAGATCCACAGA	Db	78008	AAATAATGAATAAATAAATGATGTTAAGATA	78042
repeat_region	76869 CAGCCCGGCTCACAATTTCTTATCTTATGTTTTCATGCTTTTGAAGTCTTACAGTA	QY			
repeat_region	757 ACAGGGAATTAATGCTTATTTCTTCTTTTGTAGACAGAGTCTCACTTCATCACC	Db			
repeat_region	76929 ATCATCTTTGATCTTTTTTTTTTTTGTAGACAGTCTGTCTGTCCACAGGCTGAGTGCA	QY			
repeat_region	817 TCAACCTCGGTACGTCACTGCAACCTCTGCTCCCGGTTCAAGYGAATCTCTGCT	Db			
repeat_region	76989 ATGGCGCAATCTCGGCTCACTGCAAACTCTGCTCCGAGGTTACGGGATTTCTCTGCT	QY			
repeat_region	877 AAGCTCTCTGAGTAGCTGGAATTAACAGGTCACACCACTGCTTGGCTAATTTTTT--G	Db			
repeat_region	77049 CAGCCTCAGAGTAGCTGGGATTACAGGTGGCGGCCACCACTGCTGGCTACTTTTTTTGT	QY			

RESULT 12

AP003721

LOCUS

DEFINITION

AP003721

ACCSSION

VERSION

KEYWORDS

201460 bp DNA linear
Homo sapiens genomic DNA, chromosome 11 clone:RP11-881M11, complete
sequence.
AP003721.3 GI:31790667
HTG.

SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini;
Hominidae; Homo.
1
REFERENCE Hattori,M., Ishii,K., Toyoda,A., Taylor,T.D., Hong-Seog,P.,
AUTHORS Fujiyama,A., Yada,T., Totoki,Y., Watanabe,H. and Sakaki,Y.
TITLE Homo sapiens genomic DNA
JOURNAL Published only in Database (2001)
REFERENCE Hattori,M., Ishii,K., Toyoda,A., Taylor,T.D., Hong-Seog,P.,
AUTHORS Fujiyama,A., Yada,T., Totoki,Y., Watanabe,H. and Sakaki,Y.
TITLE Direct Submission
JOURNAL Submitted (04-JUN-2001) Masahira Hattori, The Institute of Physical
and Chemical Research (RIKEN), Genomic Sciences Center (GSC);
1-7-22 Suehiro-chou,Tsukumi-ku, Yokohama, Kanagawa 230-0045, Japan
(E-mail:hattori@gsc.riken.go.jp, URL:http://hgp.gsc.riken.go.jp/
Tel:81-45-503-9111, Fax:81-45-503-9170)
COMMENT On Jun 16, 2003 this sequence version replaced gi:28189463.
FEATURES
Location/Qualifiers
source
1..201460
/organism="Homo sapiens"
/mol_type="genomic DNA"
/db_xref="taxon:9606"
/chromosome="11"
/map="11q"
/clone="RP11-881M11"
ORIGIN
Query Match 22.8%; Score 456.2; DB 8; Length 201460;
Best Local Similarity 59.7%; Pred. No. 4.8e-84;
Matches 877; Conservative 1; Mismatches 574; Indels 17; Gaps 6;
404 TTTTCTTTTGTGTTTGTAGACAGAGGCTCTGTCT- -TGTCACCCAGGCGATGACACAG 462
DB 144460 TTTCTTTTCTTTTGTAGACAGAGGTTCTCTCTGTTGCGCCAGGCTCGAGTGCA 144519
463 TGTGTCAACCATAGTCACTGCGAGCTCAACCTCTCTGAGCTCAAGGGATCTGCTGACCTC 522
DB 144520 TGCGACAATCTCGGTTTCAACCAACCTCGCGCTCCCGGGTTCAGGGTTTCTCTGCTC 144579
523 AGCCTCCCAAGTAGCTGGAGCTACGAGCGGTGACACACAGCGCTGGCTAATTAATAAAT 582
DB 144580 AGCCTCCGAGTAGCTGAGATTACAGGCGATGCGCCACCGCCAGCTGATTTGTA- - 144636
583 TTTTGTGTAGAGCTGGGCTTACTAGCTGGCGAGGCTGTCTTAACTCTCTGGCTTCA 642
DB 144637 TTATTAGTAGAGCGGGGTTTCTCCATGTTGGTCAAGCTGCTCGAGCTCTTGACCTCA 144696
643 AGCAATCCTCTACCTTGGCATCCCAAGTGTGGGATTACAGGGGTGAGCCACCATGTG 702
DB 144697 GGTGATCCACCCACTCGGCTCCCAAGTGTGGGATTATAGTGTGAGCCATCAGCC 144756
703 CGGCTACTTATTTCTTACATTCATCTTTTCCAAATAGATGTAAGATCCACAGAACAGG 762
DB 144757 CGGCTCACATATCTTATCTTAATGTTTTCACGCTTTGAAGTCTACAGTAATCATC 144816
763 ATTACTGCTATTTCTCTCTCTTTTGTGACAGAGCTCTACCTTCATCTCACTCAACC 822
DB 144817 CTTTGATTTCTTTTGTGACAGCTCTGCTCTGTCCACAGGCTGGAGTCAATGGCG 144876
823 TCGGTTTCAGTCACTGCAACTCTGCTCCCGGTTTCAAGYGATTTCTCTGCTCAAGCCT 882
DB 144877 CAATCTCGGCTCACTGCAACTCTGCTCCGAGTTTCAAGGTTTCAAGGTTTCAAGG 144936
883 CTTGAGTAGCTGGGATTACAGGCTGCGCGCCACCGCTGCTGGCTGCTGCTGCTGCTGCT 940
DB 144937 CCAGTAGCTGGGATTACAGGCTGCGCGCCACCGCTGCTGGCTGCTGCTGCTGCTGCT 144996
941 TAGCAGAGATGGGGTTTACATGTTCCCGAGGCTGCTCAACTCTGCTCAAGTCTCAAGT 1000
DB 144997 TAGTAGAGACAGGGGTTTACCATGTTGGTCAGGCTGCTGCTGCTGCTGCTGCTGCTG 145056

Qy 1001 ATCTGCTGCTCAGTCTCCAAAGTGTGGAATTTATAGCGGTGAGTCACTGTGCTGGC 1060
Db 145057 ATCGGCTGCTCAGCTTCCAAAGTGTGGAATTTATAGCGGTGAGTCACTGTGCTGGC 145116
Qy 1061 CGATTACTGTCTATTTTCTTTTATTTGCTATATCCCAAGTCTAGAGCAGTGTGCTGACAT 1120
Db 145117 GCCCTTTGATTCCTTTGAAACCAAAATGTATATAAATGCTAAACGT-TATCATTT 145175
Qy 1121 AGTAGGTGCTCAATAATAATTTGATGAATGCACAGCTAGATATAAATCTTTCTTTCTT 1180
Db 145176 CCTGTGATCATAGGCTCTGACAAATTTTCTCTAGATTTATGTTGTCAATATGATTT 145235
Qy 1181 TTTTAAACAATCTTGCACAACTTTGAGAAATTAACAATCTTGCATCTGCTTTTCA 1240
Db 145236 ATTCACACATAAAGCAAAATAATATGAGTCTACAAATTTGATGCTTTATTTATCA 145295
Qy 1241 C-----TTATCACCCTTGTATGACTTTTTCATATTTGCTCTCAAACTTTATGTTAC 1291
Db 145296 ATATAAGGATGATGACGAGAAACTAAATGTTATTTGAAAGGTATCTGAATGGTG 145355
Qy 1292 TGTTTTTCAATTTGATCTATTTTAGTCACTGAATAATATGCTTAATTTGCTTATACATC 1351
Db 145356 TTTAGGTTTTTCCCTCTCCTTACTTAACTATGATGATGATGATGATGATGATGATGATG 145415
Qy 1352 CTCTGCTCCACTTTAGAGGCCAAATTTTCAAAATCTGATGAAAGCTATGAACCTCTCC 1411
Db 145416 AACAAAGCAGAAATTTAGCATCTCTGTTTTCATTTTATGATATGATGATGATGATG 145475
Qy 1412 CCAGAGAAATACACACACACACACACACACTCACACAGT-TTTTTTTTAAATGTTTGCAC 1470
Db 145476 AGCTAGAAGGATTTGTTAGAGTGAATTTTCCCAACTCTGTCCTGAGGAGGCAAAA 145535
Qy 1471 TAAGACAAGAAACCTGCAATGAGAGATGTTTGTTCATATTAATTTAAATAAATCACTG 1530
Db 145536 TGTCAATGTAATCTAGAGTCAAAATTTATTTTCTTTTAAATGATGTTGAAGGCCA 145595
Qy 1531 GGCACAGTCACTCAAGCTGTAAACACACAGTACTTTTGAAGTCCAAGTGGGTGATCACT 1590
Db 145596 GGTGGGTGGCTCAGTATGATTAATCAAGCTTTTGAAGGCCGAGGCGGAGTGTACT 145655
Qy 1591 TGAGGTGAGAAAGTTCGAGACCAAGCTGCTCAATATGTTGAAACCTCTCTCTACTAAAA 1650
Db 145656 TGAGTTCAGAGTTCGAGACCAAGCTGCTCAATATGTTGAAACCTCTCTCTACTAAAA 145715
Qy 1651 TACAAAAATAGCTGGGTGTAGTATGATGCTGTAGTCCAGTACTCTCGGAGGCTGA 1710
Db 145716 TACAAAAATAGCTGGGTGTAGTATGATGCTGTAGTCCAGTACTCTCGGAGGCTGA 145775
Qy 1711 GGCAGAGAAATGCTTGAACCTGGGAGGCGAGGTTGAGTGGAGCGGAGATCCACCACT 1770
Db 145776 GGCAGAGAAATGCTTGAACCTGGGAGGCGGAGGTTGAGTGGAGCGGAGATCCACCACT 145835
Qy 1771 GCATCCAGCTGGGCGACACAGCAGACTCTATCTCAAAAAATAAATAAATAAATAA 1830
Db 145836 GCATCCAGCTGGGCGACACAGCAGACTCTATCTCAAAAAATAAATAAATAAATAA 145895
Qy 1831 AGGATCGGAGAGAAACAAACCTAATAAGA 1859
Db 145896 ATGAAATAATAAATAATGATGTTAAAGATA 145924
RESULT 13
AC116933
LOCUS
DEFINITION Papio anubis clone RP41-1C24, WORKING DRAFT SEQUENCE, 4 ordered
pieces.
AC116933
VERSION AC116933.2 GI:27356688
KEYWORDS HTG; HTGS_PHASE2; HTGS_DRAFT.
SOURCE Papio anubis (olive baboon)
ORGANISM Papio anubis
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;


```
Db 53494 GCTTTGGCCCTCCAAAGTGCTAGGATTACAGCATGAGCCACTGCACCGGCGCTGATCC 53435
Qy 715 TCTTTACATTCATCTTTTCCAAATAGATGTAAGATCCACAGAACAGGGAATTAAGCTGAT 774
Db 53434 TGAATTCCTTTCTTTT-----TTCCTTCT 53409
Qy 775 TTTCTTCCTTTCTTTTGGAGACAGAGTCTCATCTTCATCACTCAACCTCCGTTCA----- 830
Db 53408 TTTTCTTTTCTTTTGGAGCGAGTCTCGCTCTGTGCCCGGCTGGAGTGCAGTGG 53349
Qy 831 -----GCTCACTGCACTCTGCTCCCGGGTTCAAGYGATTTCTCTGCTCAAGC 880
Db 53348 CACGATCTTGGCTCACTGCAAGGTCACTCTCTGGGTTACGCCATTTCTCTGCTCAGC 53289
Qy 881 CTCCTGAGTAGCTGGAATTTACAAAGGTGCACCAACCATGCTTGGCTAAATTTTGTATTTT 940
Db 53298 CTCGGAGTAGCTGGGACTACAGGTGCCGCCACCACACCCAGCTAAATTTTGTATTTT 53229
Qy 941 TAGCAGAGATGGGGTTTACCAATGTTGCCAGGCTGGTCTCAAACTCTGACCTCAAGTG 1000
Db 53228 TAGTAGAGACGGGGTTTACACCGTGTAGCCAGGATGGTCTCGATCTCTGACCTC-----GTG 53171
Qy 1001 ATCTGCTGCTCAGTCTCCAAAGTCTGGAATTTATAGGCTGAGTCACTGTGCTGGC 1060
Db 53170 ATCCGCTGCTCGGCTCCCAAGTCTGGGATTTACAAAGCATGAGCCACCGGCTGGC 53111
Qy 1061 CGAATTAAGTCTATTTCTTTTATGCTATATCCCAAGTCTAGAGCAGTGTGCAATAT 1120
Db 53110 CTTTCTTTTATTTTGTGACAGAGGCTTC-CCTCTGTCAACCGAGTGTAAATGC 53052
Qy 1121 AGTAGGTGCTCAATAAATATGATGAATGCACAGCTAGATATAAATCTTCTTTCTT 1180
Db 53051 ACTGGCACTATCTGGCTCACTGCAACCTCCACTTCCAGGTTCAAGCAATTAATGCGC 52992
Qy 1181 TTTTAAACAATCTTGCAACCTTTGCAAAATTAACAATCTTGCAATCTGCTTTTCA 1240
Db 52991 TCAGTCTCCTGATACTGGGATACAGGGGTGCGCGTACGCCAGCTAATTTTGTAT 52932
Qy 1241 CTTATCACTTGTATGACTTTTTTCAATTTGCTCAAAAC---CTTTATTTGTTACTGTTTT 1297
Db 52931 TTTTGTAGTAAGAGGAGTTTCCCAATGTGGCAAGATGGTCTCCATCTCTGACCTCG 52872
Qy 1298 TTCATTTGTTACTATTTTAGTCACTGAATAATATGCGCTTAATTTGCTTATACATCTCCCTG 1357
Db 52871 TGATCTGCCGCTCGGCTCCCAAGTGTGAGATTACAGGCGTGAGCACTGTGCGCG 52812
Qy 1358 CTCACCTTTAGAGGCCAAATTTACAAATCTGATGAAGCTATGAACCTCTCCCCAGAG 1417
Db 52811 GCCTGATAACAGCTTTCTATGTGCCAGCTACTGCTCCAAAGTTTATGTTATATATGAT 52752
Qy 1418 AATATACACACACACACACTCACACAGTTTTTTTTTAAATGTTTGCAACTAAGACA 1477
Db 52751 TAATTCATTAGTCAATTTGTTATTTATATGTAATTAATTCCTTTTATAGTCCATTTTC 52692
Qy 1478 AGAAACCTGCATTAGAGGATGTTTGTTCATATTAATTAATAAATAAATCACTAGTTGGGCAG 1537
Db 52691 ACCCTGCTGATAAAGACATATCTTGAGACTGAGAAGAAAGATGTTTAGGCCAGGAGTGG 52632
Qy 1538 TGACTCAGCTGTATACACAGTACTTTGGAAGTCCAAAGTGGGTGATCACTTGAAGTGG 1597
Db 52631 TAGCTCACACCTGTAATCCAGCACTTTGGAGGCGGAGGTGGAAGGATTAACCTGAGATC 52572
Qy 1598 AGAAGTTCGAGACCGAGCTGGTCAATATGTTGAAACCTTATCTCTACTAAAAATACAAAA 1657
Db 52571 AGAGTTTGAGACCACTCGGCCAACATGCAAAACCCCATCTCTACTAABAAATACAAA 52512
Qy 1658 ATTAGCTGGGTAGTATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1717
Db 52511 ATTAGTAGGGGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 52452
Qy 1718 GAATTCCTGAACTGGGAGGAGGTTGAGTGGAGCGGAGATCCCAACCACTGCACTCC 1777
```

```
Db 52451 GAATTCCTGAACTGGGAGGTTGAGTGGAGGATCAGCCAGGATCAGCCACTGCACTCC 52392
Qy 1778 AGCTGGGCGACAGGAGGAGTCTATCTCAAAAAATAAATAAATAAATAAATAAATAAATAA 1837
Db 52391 AGCTGGGTGACAGATGAGACTCCATTAATAAAAAAAAAAAAAAAAAAGAAAGAA 52332
Qy 1838 GAGAGA 1843
Db 52331 AAAAGA 52326

RESULT 15
AC005484/c
LOCUS Homo sapiens PAC clone RP5-84708 from 14q24.3, complete sequence.
DEFINITION AC005484
VERSION AC005484.2 GI:5091654
KEYWORDS HTG.
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini;
Hominidae; Homo.
REFERENCE 1 (bases 1 to 131943)
AUTHORS Sulston, J.E. and Waterston, R.
TITLE Toward a complete human genome sequence
JOURNAL Genome Res. 8 (11), 1097-1108 (1998)
PUBMED 9847074
REFERENCE 2 (bases 1 to 131943)
AUTHORS Clout, J., Wohlmann, P. and Holmes, A.
TITLE The sequence of Homo sapiens PAC clone RP5-84708
JOURNAL Unpublished
REFERENCE 3 (bases 1 to 131943)
AUTHORS Waterston, R.H.
TITLE Direct Submission
JOURNAL Submitted (14-AUG-1998) Genome Sequencing Center, Washington
University School of Medicine, 4444 Forest Park Parkway, St. Louis,
MO 63108, USA
REFERENCE 4 (bases 1 to 131943)
AUTHORS Waterston, R.H.
TITLE Direct Submission
JOURNAL Submitted (17-JUN-1999) Genome Sequencing Center, Washington
University School of Medicine, 4444 Forest Park Parkway, St. Louis,
MO 63108, USA
REFERENCE 5 (bases 1 to 131943)
AUTHORS Waterston, R.
TITLE Direct Submission
JOURNAL Submitted (30-SEP-2000) Department of Genetics, Washington
University, 4444 Forest Park Avenue, St. Louis, Missouri 63108, USA
On Jun 17, 1999 this sequence version replaced gi:3907511.
COMMENT
----- Genome Center
Center: Washington University Genome Sequencing Center
Center code: WUGSC
Web site: http://genome.wustl.edu/gsc
Contact: sapiens@wustl.wustl.edu
----- Summary Statistics
-----
Center project name: H_DJ0847008
-----

NOTICE: This sequence may not represent the entire insert of this
clone. It may be shorter because we only sequence overlapping
clone sections once, or longer because we provide a small overlap
between neighboring data submissions.

This sequence was finished as follows unless otherwise noted:
all regions were double stranded, sequenced with an alternate
chemistry, or covered by high quality data (i.e., phred quality >=
30); an attempt was made to resolve all sequencing problems, such
as compressions and repeats; all regions were covered by sequence
from more than one subclone; and the assembly was confirmed by
restriction digest.

MAPPING INFORMATION:
```

This clone from chromosome 14 was provided by Dr. Pieter de Jong, Roswell Park Cancer Institute, Human Genetics Department, Elm and Carlton Streets, Buffalo NY 14263-0001 USA.

SOURCE INFORMATION:

This clone was derived from human PAC library RPCI-5, prepared by Pieter de Jong and coworkers at the Roswell Park Cancer Institute (<http://bacpac.med.buffalo.edu>) using the method described by Ioannou et al., Nature Genetics 6:84-9 (1994). The library is from one male donor.

The clone may be obtained either from Genome Systems, Inc. (<http://www.genomesystems.com>) or Research Genetics, Inc. (<http://www.resgen.com>); or from Pieter de Jong.

VECTOR: pCYPAC2

NEIGHBORING SEQUENCE INFORMATION:

The clone sequenced to the left is RP5-892G5; the clone sequenced to the right is RP4-592G7, 200 bp overlap. Actual start of this clone is at base position 1 of RP5-84708; actual end is at base position 131747 of RP5-84708.

FEATURES

source

```
1..131943
/organism="Homo sapiens"
/mol_type="genomic DNA"
/db_xref="taxon:9606"
/chromosome="14"
/map="14q24.3"
/clone="RP5-84708"
/clone_lib="RPCI-5"
1058..1098
/rpt_family="AT_rich"
repeat_region
1762..2225
/rpt_family="MaLR"
repeat_region
2305..2587
/rpt_family="Alu"
repeat_region
2937..3086
/rpt_family="L2"
repeat_region
3293..3591
/rpt_family="Alu"
repeat_region
3651..3916
/rpt_family="L1"
repeat_region
3917..3943
/rpt_family="(TA)n"
repeat_region
4815..4841
/rpt_family="AT_rich"
repeat_region
5153..5244
/rpt_family="L1"
repeat_region
5728..6008
/rpt_family="Alu"
repeat_region
6009..6034
/rpt_family="(CAA)n"
repeat_region
8080..8468
/note="match to EST R05271 (NID:g7555891) ye91b03.s1"
misc_feature
8080..8277
STS
8284..8589
/rpt_family="Alu"
repeat_region
8649..9039
/note="match to EST R05328 (NID:g755948) ye91b03.r1"
misc_feature
8657..9007
/rpt_family="MIR"
repeat_region
9062..9152
/rpt_family="MIR"
repeat_region
9216..9288
/rpt_family="MIR"
repeat_region
9476..9770
/rpt_family="Alu"
repeat_region
9711..10034
STS
10004..10114
/rpt_family="Alu"
repeat_region
10115..10135
/rpt_family="(TAAA)n"
repeat_region
10136..10158
/rpt_family="Alu"
repeat_region
10205..10501
```

```
repeat_region
10513..10819
/rpt_family="Alu"
repeat_region
10890..11242
/rpt_family="MaLR"
repeat_region
11246..11519
/rpt_family="Alu"
repeat_region
11567..11996
/rpt_family="L2"
repeat_region
12006..12113
/rpt_family="L2"
repeat_region
12123..12146
/rpt_family="(CAAAA)n"
repeat_region
12147..12430
/rpt_family="Alu"
repeat_region
12445..12535
/rpt_family="L1"
repeat_region
12527..13097
/rpt_family="L1"
repeat_region
13217..13339
/rpt_family="Alu"
repeat_region
13389..13994
/rpt_family="L2"
repeat_region
14138..14158
/rpt_family="AT_rich"
repeat_region
14195..15071
/rpt_family="L1"
repeat_region
15048..15521
/rpt_family="L1"
repeat_region
15520..15992
/rpt_family="L1"
repeat_region
15989..16221
/rpt_family="L1"
repeat_region
16257..16556
/rpt_family="Alu"
repeat_region
16569..16963
/rpt_family="L1"
repeat_region
17016..17039
/rpt_family="AT_rich"
repeat_region
17073..17182
/rpt_family="MIR"
repeat_region
17252..17452
/rpt_family="MaLR"
repeat_region
17473..17991
/rpt_family="L2"
repeat_region
18106..18192
/rpt_family="MERI_type"
repeat_region
18193..18504
/rpt_family="Alu"
repeat_region
18505..18588
/rpt_family="MERI_type"
repeat_region
18624..18904
/rpt_family="L1"
repeat_region
19144..19368
/rpt_family="Alu"
repeat_region
20030..20159
/rpt_family="Alu"
repeat_region
20167..20317
/rpt_family="Alu"
repeat_region
20318..20617
/rpt_family="Alu"
repeat_region
20618..20753
/rpt_family="Alu"
repeat_region
20802..20954
/rpt_family="L1"
repeat_region
20955..21256
/rpt_family="Alu"
```

Query Match 22.4%; Score 447.8; DB 8; Length 131943;
Best Local Similarity 61.3%; Pred. No. 2.7e-82;
Matches 913; Conservative 1; Mismatches 523; Indels 52; Gaps 10;

Qy

397 GTTTGGGTTTTTTTTTTTGTGTTGTTTAGACAGAGGGCTTTGCTC-TGTCACCCAGGCATG 455

[illegible]

Search completed: February 9, 2006, 18:20:35
Job time : 10781.2 secs

GenCore version 5.1.1.7
Copyright (c) 1993 - 2006 Bioceleration Ltd.

OM nucleic - nucleic search, using sw model

Run on: February 9, 2006, 13:16:14 ; Search time 1377.77 Seconds
(without alignments)
9679.442 Million cell updates/sec

Title: US-10-607-806-1_COPY_7200_9200

Perfect score: 1998.6

Sequence: 1 ctatctcgctgtcctctccc.....ckctacataaaatatttatta 2001

Scoring table: IDENTITY NUC

Gapop 10_0 , Gapext 1.0

Searched: 4996997 seqs, 332346308 residues

Total number of hits satisfying chosen parameters: 9993994

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : N_Geneseq_21:*

- 1: Geneseqn1980s:*
- 2: Geneseqn1990s:*
- 3: Geneseqn2000s:*
- 4: Geneseqn2001as:*
- 5: Geneseqn2001bs:*
- 6: Geneseqn2002as:*
- 7: Geneseqn2002bs:*
- 8: Geneseqn2003as:*
- 9: Geneseqn2003bs:*
- 10: Geneseqn2003cs:*
- 11: Geneseqn2003ds:*
- 12: Geneseqn2004as:*
- 13: Geneseqn2004bs:*
- 14: Geneseqn2005s:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	1998.6	100.0	12174	12 ADI35082	Adi35082 Human PLA
2	1998.6	100.0	12174	12 ADJ09983	Adj09983 Human pho
3	1998.6	100.0	13612	6 ABK47376	Abk47376 Human pho
4	434.2	21.7	93544	13 ABD33504	Abd33504 Human can
5	434.2	21.7	160482	11 ACN43914	Acn43914 Human gen
6	433.2	21.7	91760	11 ACN44410	Acn44410 Human gen
7	425.4	21.3	38538	13 ABD33357	Abd33357 Human can
8	417.2	20.9	122888	6 ABK83569	Abk83569 Human cDN
9	416.2	20.8	6405	5 AAF97850	Aaf97850 Human neu
10	416	20.8	126990	12 ADP13332	Adp13332 Renal cel
11	414.2	20.7	196686	11 ACN44170	Acn44170 Human gen
12	413.6	20.7	86000	12 ADP68568	Adp68568 Human PPA
13	413.6	20.7	215974	12 ADQ97523	Adq97523 Human can
14	413.4	20.7	13670	6 AAL42369	Aal42369 Human Gua
15	413.2	20.7	6519	5 ABA16909	Abal6909 Human ner
16	408.8	20.5	17245	4 AAK83897	Aak83897 Human imm
17	405.2	20.3	348101	12 ADQ97146	Adq97146 Human can
18	405.8	20.3	227246	13 ABD33272	Abd33272 Human can
19	404	20.2	95240	10 ADL13556	Adl13556 Osteoarth

c 20	403.6	20.2	23639	6 ABQ78991	Abq78991 Human pho
c 21	400.4	20.0	14176	4 AAS26670	Aas26670 Human gen
c 22	400.4	20.0	14176	8 ABX74019	Abx74019 Human nov
c 23	400.4	20.0	177531	8 ACF62732	Acf62732 Cancer ba
c 24	400.4	20.0	177531	8 ADB20847	Adb20847 MRP1 base
c 25	400.4	20.0	177531	10 ADB87936	Adb87936 Human UGT
c 26	400.4	20.0	177531	10 ADB96919	Adb96919 Human MDR
c 27	400.4	20.0	177531	10 ADB92110	Adb92110 Human MDR
c 28	400.4	20.0	177531	10 ADH74617	Ach74617 Human BAC
c 29	399	20.0	87687	11 ACN45166	Acn45166 Human gen
c 30	399	20.0	131078	14 ADX06911	Adx06911 Cyclin-de
c 31	398.6	19.9	174448	11 ACN43946	Acn43946 Human gen
c 32	398.4	19.9	167343	6 ABL64403	Abi64403 Stomach c
c 33	398.4	19.9	167343	6 ABL67239	Abi67239 Thyroid c
c 34	394.8	19.8	30620	4 AAK66931	Aak66931 Human imm
c 35	394.6	19.7	19300	12 ADP74371	Adp74371 Human X c
c 36	392.4	19.6	9289	10 ADE84033	Ade84033 5' regula
c 37	391.4	19.6	226475	9 AAD58279	Aad58279 Human tum
c 38	390	19.5	13224	4 AAS41751	Aas41751 Genomic s
c 39	390	19.5	13224	4 ABA06811	Abao6811 Human gen
c 40	390	19.5	13224	6 ABV84148	Abv84148 Human pol
c 41	390	19.5	36221	4 AAS00624	Aas00624 Human dea
c 42	389.6	19.5	23456	13 ABD33110	Abd33110 Human can
c 43	389.6	19.5	60057	11 ACN44314	Acn44314 Human gen
c 44	388	19.4	75252	11 ACN44450	Acn44450 Human gen
c 45	387.8	19.4	48001	12 ADO61194	Ado61194 Human PPA

ALIGNMENTS

RESULT 1

ADI35082
ID ADI35082 standard; DNA; 12174 BP.

XX
AC ADI35082;

XX
DT 22-APR-2004 (first entry)

XX
DE Human PLA2G1B nucleotide sequence.

XX
KW PLA2G1B ; fat deposition; leanness; polymorphism;

XX
KW non-insulin dependent diabetes mellitus; NIDDM; hyperinsulinemia;

XX
KW hypertension; Glucose intolerance; dyslipidemia; hypercoagulability;

XX
KW microalbuminuria; human; gene; ds.

XX
OS Homo sapiens.

XX
FN WO2004002295-A2.

XX
XX 08-JAN-2004.

XX
PD 27-JUN-2003; 2003WO-US020830.

XX
PF 27-JUN-2002; 2002US-0392361P.

XX
PR 27-JUN-2002; 2002US-0392361P.

XX
PA (SEQU-) SEQUENOM INC.

XX
PI Adam GIR, Langdown ML;

XX
DR WPI; 2004-082843/08.

XX
DR P-PSDB; ADI35083.

XX
PT Diagnosing a predisposition to fat deposition or leanness, useful for
PT diagnosing a predisposition to e.g. diabetes or hypertension, comprises
PT detecting the presence of a polymorphism in the PLA2G1B nucleic acid from
PT the subject.

XX
PS Claim 1; SEQ ID NO 1; 91pp; English.

XX
CC The invention relates to diagnosing a predisposition to fat deposition or
CC leanness in a subject comprising detecting the presence or absence of a
CC polymorphic variation associated with fat deposition at a polymorphic

site in a PLA2G1B nucleotide sequence in a nucleic acid sample from a subject, where the presence of the polymorphic variation indicates a predisposition to fat deposition in the subject. The polymorphic variation is a guanine at position 7328 or thymine at position 9182 of the present sequence. The method is useful for diagnosing a predisposition to fat deposition or leanness in a subject, and consequently for diagnosing a predisposition to non-insulin dependent diabetes mellitus (NIDDM) in a subject and conditions such as hyperinsulinemia, hypertension, glucose intolerance, dyslipidemia, hypercoagulability, or microalbuminuria, which can lead to early prescription of preventive measures. The present sequence represents a human PLA2G1B nucleotide sequence.

xx	sq	Sequence	12174 BP; 3217 A; 2992 C; 2738 G; 3215 T; 0 U; 12 Other;
		Query Match	100.0%; Score 1998.6; DB 12; Length 12174;
		Best Local Similarity	100.0%; Pred. No. 0;
		Matches 2001; Conservative	0; Mismatches 0; Indels 0; Gaps 0;
Qy	1	CTATCTCAGCTGTCCTCCCACTTTCAGGTGTCGACACATGACAACTGCTAYGAC	60
Db	7200	CTATCTCAGCTGTCCTCCCACTTTCAGGTGTCGACACATGACAACTGCTAYGAC	7259
Qy	61	CAGGCCAAGAGCTGACAGCTGTAATTTCTGCTGACAMCGGTACACCACTAT	120
Db	7260	CAGGCCAAGAGCTGACAGCTGTAATTTCTGCTGACAMCGGTACACCACTAT	7319
Qy	121	TCATACTCRCTGCTCTGGCTCGGCAATCACCTGTAGCTAGGTTTATCCCTTCCTTGACC	180
Db	7320	TCATACTCRCTGCTCTGGCTCGGCAATCACCTGTAGCTAGGTTTATCCCTTCCTTGACC	7379
Qy	181	TATGAATTTCTAGTTGGTTCTCAGTAGCGCGGGGGGAAATAATAGTAAACACGCCATGAT	240
Db	7380	TATGAATTTCTAGTTGGTTCTCAGTAGCGCGGGGGGAAATAATAGTAAACACGCCATGAT	7439
Qy	241	TTAGTGTAAATTTTCTGGTCTCGGCAGTGTCTCTTTAATCCTCAGAAACAACTATG	300
Db	7440	TTAGTGTAAATTTTCTGGTCTCGGCAGTGTCTCTTTAATCCTCAGAAACAACTATG	7499
Qy	301	GGATAGGTACAATTATCCTCACTTAAACAGATAAGAAAACCTGAGCCTCAGAAGGCTGAGCT	360
Db	7500	GGATAGGTACAATTATCCTCACTTAAACAGATAAGAAAACCTGAGCCTCAGAAGGCTGAGCT	7559
Qy	361	ATTGCGCAAGATCAACAGCTTGTAGTGTGTACAGTTTGGGTTTTTTTTTGTGTGTGT	420
Db	7560	ATTGCGCAAGATCAACAGCTTGTAGTGTGTACAGTTTGGGTTTTTTTTTGTGTGTGT	7619
Qy	421	TTAGAGACAGGGTCTTGCTCTGTCTACCCAGGCATGAGCACAGTGGTGCACCATAGGTCA	480
Db	7620	TTAGAGACAGGGTCTTGCTCTGTCTACCCAGGCATGAGCACAGTGGTGCACCATAGGTCA	7679
Qy	481	CTGCAGCCTCAAACCTCCTGAGCTCAAGGATCTGCTGACCTCCCAAGTAGCTGG	540
Db	7680	CTGCAGCCTCAAACCTCCTGAGCTCAAGGATCTGCTGACCTCCCAAGTAGCTGG	7739
Qy	541	GACTACGAGCTGACACCAACCGCTGGCTAAATTAABAAAATTTTTTGTAGAGACTGGG	600
Db	7740	GACTACGAGCTGACACCAACCGCTGGCTAAATTAABAAAATTTTTTGTAGAGACTGGG	7799
Qy	601	TCTTACTAGCTTTGGCCAGGCTTGTCTTAACTCCTGGCTTCAAGCAATCCTCCTACCTTG	660
Db	7800	TCTTACTAGCTTTGGCCAGGCTTGTCTTAACTCCTGGCTTCAAGCAATCCTCCTACCTTG	7859
Qy	661	GCATCCAAAGTGTGGGATTAACGGGTGACCCACCATCTCGCGCTACTTATTTCTTTA	720
Db	7860	GCATCCAAAGTGTGGGATTAACGGGTGACCCACCATCTCGCGCTACTTATTTCTTTA	7919
Qy	721	CATTCCATCTTTCCCAATAGATGTAAGTCCACAGAACAGGGATTACTGCCTATTTTCTT	780
Db	7920	CATTCCATCTTTCCCAATAGATGTAAGTCCACAGAACAGGGATTACTGCCTATTTTCTT	7979
Qy	781	CCTTTCTTTTTTGAGACAGAGTCTCATTTCACTCAACTCCGTTGAGTCACTGCA	840

QY 1921 GTAATCTTATTTGTTTATTTGTTTATTAAGTAACAAGCCAAAAGTAATGCAACTTCAA 1980
DB |||||||
QY 9120 GTAATCTTATTTGTTTATTTGTTTATTAAGTAACAAGCCAAAAGTAATGCAACTTCAA 9179
DB |||||||

QY 1981 ACKCTACATAATATCTATTA 2001
DB |||||||

QY 9180 ACKTACATAATATCTATTA 9200
DB |||||||

RESULT 2

ADJ09983

ID ADJ09983 standard; DNA; 12174 BP.

XX AC ADJ09983;

XX DT 17-JUN-2004 (first entry)

XX DE Human phospholipase A2 (PLA2G1B) DNA SeqID 1.

XX KW human; gene; ds; fat reduction; fat deposition; phospholipase A2;

KW PLA2G1B; chromosome 12q24; single nucleotide polymorphism; SNP;

KW appetite suppressant; lipase inhibitor; exercise regimen; obesity;

KW non-insulin dependent diabetes mellitus; NIDDM; cardiovascular disorder;

KW hypertension; antidiabetic.

XX OS Homo sapiens.

XX PH Key Location/Qualifiers

FT variation replace(436,c)

FT /tag= a

FT /standard_name= "Single nucleotide polymorphisms"

FT replace(839,a)

FT /tag= b

FT /standard_name= "Single nucleotide polymorphisms"

FT replace(4050,a)

FT /tag= c

FT /standard_name= "Single nucleotide polymorphisms"

FT replace(4689,t)

FT /tag= d

FT /standard_name= "Single nucleotide polymorphisms"

FT replace(6282,a)

FT /tag= e

FT /standard_name= "Single nucleotide polymorphisms"

FT replace(6358,c)

FT /tag= f

FT /standard_name= "Single nucleotide polymorphisms"

FT replace(6653,t)

FT /tag= g

FT /standard_name= "Single nucleotide polymorphisms"

FT replace(7300,a)

FT /tag= h

FT /standard_name= "Single nucleotide polymorphisms"

FT replace(7301,c)

FT /tag= i

FT /standard_name= "Single nucleotide polymorphisms"

FT replace(7328,g)

FT /tag= j

FT /standard_name= "Single nucleotide polymorphisms"

FT replace(8062,c)

FT /tag= k

FT /standard_name= "Single nucleotide polymorphisms"

FT replace(9182,t)

FT /tag= l

FT /standard_name= "Single nucleotide polymorphisms"

FT replace(11649,c)

FT /tag= m

FT /standard_name= "Single nucleotide polymorphisms"

FT replace(11649,c)

FT /tag= n

FT /standard_name= "Single nucleotide polymorphisms"

XX PN W02004002296-A2.

XX 08-JAN-2004.
PD 27-JUN-2003; 2003WO-US020831.
XX PF 27-JUN-2002; 2002US-0392362P.
XX PR (SEQU-) SEQUENOM INC.
XX PA Adam GIR, Langdown ML, Denissenko MF, Dennis E, Cantor C;
XX PI Rubin B;
XX PI WPI; 2004-071944/07.
DR P-ESDB; ADJ09984.
XX DR
XX PT Identifying a candidate therapeutic for fat reduction, useful for
XX PT treating diabetes, by introducing a test molecule to a system comprising
XX PT PLA2G1B protein or nucleic acid, and determining the presence of
XX PT interaction between the compounds.
XX PS Claim 1; SEQ ID NO 1; 116pp; English.
XX CC This invention relates to a novel candidate therapeutic agent useful for
XX CC fat reduction and disorders related to fat depositions. Specifically, it
XX CC refers to polymorphic variations in the phospholipase A2 (PLA2G1B) DNA,
XX CC which is located on chromosome 12q24 and has been associated with central
XX CC fat deposition. The present invention describes methods to detect the
XX CC presence or absence of these single nucleotide polymorphisms of PLA2G1B,
XX CC in particular G7328A and T9182G, and subsequently provide treatment that
XX CC reduces fat deposition. This treatment may consist of an appetite
XX CC suppressant, a lipase inhibitor, a phospholipase inhibitor, an exercise
XX CC regimen, a dietary regimen, psychological counselling, psychotherapy or a
XX CC psychotherapeutic. Accordingly, PLA2G1B is a target for reducing fat
XX CC deposition and it can be used to treat both obesity and non-insulin
XX CC dependent diabetes mellitus (NIDDM), as well as cardiovascular disorders
XX CC such as hypertension. As such, it exhibits antidiabetic activity. This
XX CC polynucleotide sequence is the human PLA2G1B DNA of the invention.
XX SQ Sequence 12174 BP; 3220 A; 2996 C; 2739 G; 3219 T; 0 U; 0 Other;

Query Match 100.0%; Score 1998.6; DB 12; Length 12174;
Best Local Similarity 99.7%; Pred. No. 0;
Matches 1995; Conservative 6; Mismatches 0; Indels 0; Gaps 0;
Qy 1 CTATCTCAGCTGCTCCCTCCACATTTCCAGGTGCTGCAGACATCATCACTGCTATGAC 60
Db 7200 CTATCTCAGCTGCTCCCTCCACATTTCCAGGTGCTGCAGACATCATCACTGCTATGAC 7259
Qy 61 CAGGCCAAGAAGCTGGACAGCTGTAATTTCTCTGGAACAMGCCGTACACCCACACCTAT 120
Db 7260 CAGGCCAAGAAGCTGGACAGCTGTAATTTCTCTGGAACACCGGTACACCCACACCTAT 7319
Qy 121 TCATCTCTGCTCTGGCTCGGCAATCACTGTAGCAGTAGGTTTATTCCTTCTTGACC 180
Db 7320 TCATCTCTGCTCTGGCTCGGCAATCACTGTAGCAGTAGGTTTATTCCTTCTTGACC 7379
Qy 181 TATGAATTCCTAGTTGGTCTCTCAGTAGCCGGGGGAAATAATAGTAACAACACCATGAT 240
Db 7380 TATGAATTCCTAGTTGGTCTCTCAGTAGCCGGGGGAAATAATAGTAACAACACCATGAT 7439
Qy 241 TTAGTGTAAATTTCTTGGTCTTGGCAGGTGCTCTCTTAAATCCTCAGAACACACTATG 300
Db 7440 TTAGTGTAAATTTCTTGGTCTTGGCAGGTGCTCTCTTAAATCCTCAGAACACACTATG 7499
Qy 301 GGATAGGTACAAATTATCTCTCACTTAACAGATAAGAAACCTGAGGCTCAGAAGGCTGAGCT 360
Db 7500 GGATAGGTACAAATTATCTCTCACTTAACAGATAAGAAACCTGAGGCTCAGAAGGCTGAGCT 7559
Qy 361 ATTTGGCCCAAGATCACACAGCTTGTAAAGTGGTGACAGTTGGGTTTTTTTGTGTTGT 420
Db 7560 ATTTGGCCCAAGATCACACAGCTTGTAAAGTGGTGACAGTTGGGTTTTTTTGTGTTGT 7619
Qy 421 TTAGACAGAGGGTCTTGGTCTCTGTCCACCCAGGCATGAGCAGTGGTGCAACCATAGTCA 480

Db 7676 CATTTGCCAGGCTGCTCAAACTCGTCAAGTGAATCGCTGCTCAGTCTCC 7735
Qy 1021 CAAAGTGTGGAATATATAGGCGTGAGTCACTGTGCTGGCGGATCTACTGTCTATTTTCTT 1080
Db 7736 CAAAGTGTGGAATATATAGGCGTGAGTCACTGTGCTGGCGGATCTACTGTCTATTTTCTT 7795
Qy 1081 TATTGCTATATCCCCAGATCTAGAGCAGTGTCTGCATATATAGTAGTGCTCAATAAATAA 1140
Db 7796 TATTGCTATATCCCCAGATCTAGAGCAGTGTCTGCATATATAGTAGTGCTCAATAAATAA 7855
Qy 1141 TTGATGAATGCACAGCCTAGATATAAACTTTCTTTTCTTTTCTTTTAAACAATCTTGACA 1200
Db 7856 TTGATGAATGCACAGCCTAGATATAAACTTTCTTTTCTTTTCTTTTAAACAATCTTGACA 7915
Qy 1201 ACTTTGCAGATAAATCAATCTTGCAATCTGCTTTTCTTTTCTTTTCACTATACCTTGTATGACT 1260
Db 7916 ACTTTGCAGATAAATCAATCTTGCAATCTGCTTTTCTTTTCTTTTCACTATACCTTGTATGACT 7975
Qy 1261 TTTTCATATTCGCTCAAACTTTTATTTGTTACTGTTTCTTTTCTTTTCAATGTTACTATTTAGTCAC 1320
Db 7976 TTTTCATATTCGCTCAAACTTTTATTTGTTACTGTTTCTTTTCTTTTCAATGTTACTATTTAGTCAC 8035
Qy 1321 TGAATAATATGGCTTAATTTGCTTATACATCCTCTGCTGCTCCACTTTAGAGGCGCAATTT 1380
Db 8036 TGAATAATATGGCTTAATTTGCTTATACATCCTCTGCTGCTCCACTTTAGAGGCGCAATTT 8095
Qy 1381 ACAATCTGATGAAGCTATGAACCTCTCCAGAGAAATACACACACACACACT 1440
Db 8096 ACAATCTGATGAAGCTATGAACCTCTCCAGAGAAATACACACACACACACT 8155
Qy 1441 CACACACAGTTTCTTTTAAATGTTTGAACCTTAAGCAAGAACTGCAATTAGAGGATGTT 1500
Db 8156 CACACACAGTTTCTTTTAAATGTTTGAACCTTAAGCAAGAACTGCAATTAGAGGATGTT 8215
Qy 1501 TGTTCATATTAATTAATAAATACTCAGTTGGGCAAGTGAATCAAGCTGTAAACACAGT 1560
Db 8216 TGTTCATATTAATTAATAAATACTCAGTTGGGCAAGTGAATCAAGCTGTAAACACAGT 8275
Qy 1561 ACTTTGGAAGTCAAGGTGGTGGATCACTTGAGGTGAGAACTCGAGACAGCCTGGTC 1620
Db 8276 ACTTTGGAAGTCAAGGTGGTGGATCACTTGAGGTGAGAACTCGAGACAGCCTGGTC 8335
Qy 1621 AATATGTCGAACCTTATCTCTACTAAATAATCAAAAAATTAAGTCTGGGTGTAGTGCAT 1680
Db 8336 AATATGTCGAACCTTATCTCTACTAAATAATCAAAAAATTAAGTCTGGGTGTAGTGCAT 8395
Qy 1681 GCCTGTAGTCCAGCTACTCGGGAGGCTGAGGCAAGAGAAATGCTTGAACCTGGAGGCA 1740
Db 8396 GCCTGTAGTCCAGCTACTCGGGAGGCTGAGGCAAGAGAAATGCTTGAACCTGGAGGCA 8455
Qy 1741 GAGTTGCAAGTGGCGAGATCCACCACTGCACTCCAGCCTGGGCGACACAGCGAGCT 1800
Db 8456 GAGTTGCAAGTGGCGAGATCCACCACTGCACTCCAGCCTGGGCGACACAGCGAGCT 8515
Qy 1801 CTATCTCAAAAAATAATAATAATAATAAGATCGAGAGAAAAAATAAATAAAGAT 1860
Db 8516 CTATCTCAAAAAATAATAATAATAATAAGATCGAGAGAAAAAATAAATAAAGAT 8575
Qy 1861 TCCTGAAGTGAAGCAGATACGTAAATATATATATATATATATATATATATATATATATAT 1920
Db 8576 TCCTGAAGTGAAGCAGATACGTAAATATATATATATATATATATATATATATATATATAT 8635
Qy 1921 GTAATCTTATTTGTTTATTTGTTTATTAAGTATAAAGTAAAGTAAAGTAAAGTAAAGT 1980
Db 8636 GTAATCTTATTTGTTTATTTGTTTATTAAGTATAAAGTAAAGTAAAGTAAAGTAAAGT 8695
Qy 1981 ACKTACATAAATATCTATTA 2001
Db 8696 ACTCTACATAAATATCTATTA 8716

ABD33504

ID ABD33504 standard; DNA; 93544 BP.

XX AC ABD33504;

XX DT 18-NOV-2004 (first entry)

XX DE Human cancer-associated (CA) gene HD07-098.

XX DX Human; cancer-associated protein; CAP; cancer-associated gene; CA; gene;
KW ds; cancer; cytostatic.

XX OS Homo sapiens.

XX PN WO2004059146-A2.

XX PD 15-JUL-2004.

XX PF 15-DEC-2003; 2003WO-US040081.

XX PR 17-DEC-2002; 2002US-00322281.

XX PA (SAGR-) SAGRES DISCOVERY INC.

XX PI Morris DW, Malandro MS;

XX DR WPI; 2004-499109/47.

XX PT Novel human cancer associated protein encoded within open reading frame
XX of cancer associated gene, useful as targets for diagnosing cancer.

XX PS Claim 16; SEQ ID NO 676; 182pp; English.

XX CC The invention relates to cancer-associated proteins (CAP) and the cancer-
associated (CA) nucleic acids encoding them. The invention also relates
to a method for treating cancers involving administering to a patient an
inhibitor of CAP, and a method of screening for anticancer activity in a
potential drug involving providing a cell that expresses a CA gene,
contacting a tissue sample derived from a cancer cell with an anticancer
drug candidate and monitoring the effect of the anticancer drug candidate
on expression of the CA gene. The CAP proteins are useful for detecting
cancer associated with expression of a CAP protein in a test cell sample
and for screening for a bioactive agent capable of modulating the
activity of a CAP protein. The CA nucleic acids are useful for diagnosing
cancer involving determining the expression of a CA nucleic acid in a
tissue. This sequence represents a human CA gene of the invention. Note:
The sequence data for this patent did not form part of the printed
CC specification, but was obtained in electronic format directly from WIPO
at ftp.wipo.int/pub/published_pct_sequences

XX SQ Sequence 93544 BP; 26138 A; 20027 C; 21635 G; 25744 T; 0 U; 0 Other;

Query Match 21.7%; Score 434.2; DB 13; Length 93544;

Best Local Similarity 59.7%; Pred. No. 5.3e-82;

Matches 861; Conservative 1; Mismatches 559; Indels 22; Gaps 7;

Qy 404 TTTT TTTT TTTT TTTT TTTT TAGAGACAGGCTTGTCTGTCTACCCAGCATGAGCAGCT 463

Db 18099 TTTT TTTT TTTT TTTT TTTT TAGAGATGGAGTCTTAGCTGTGCCAGGCTGGAGTGCAT 18158

Qy 464 GGTGCAACCATAGTCACTGAGCCTCAACCTCTGAGCTCAAGGATCTGCTGACCTCA 523

Db 18159 GGCCACCATCTGGGTCACTGCAACCTCCACATGTGGGTTCAAGAGATTTCTACTGCTCA 18218

Qy 524 GCCTCCCAAGTAGCTGGGACTACGAGCGTGCA CCAACGCGCTGGCTAATTAATAAATTT 583

Db 18219 ACCTCCCAAGTAGCTGGGATTAACAGCTTGTGCCACCAACACCTTGTGTAATTT 18276

Qy 584 TTTT TTTT TAGAGACTGGTCTTACTAGTGTGGCAGGCTGTCTTAACTCCTGGCTTCAA 643

Db 18277 TTTT TTTT TAGAGACAGGGTTTCGCCATGTGTGGCAGGCTGATCTCAAACTCTGGGCTCAA 18336

Qy 644 GCAATCCTCTCTACCTTGGCATCCCAAGTGTCTGGGATTACAGGGGTGAGCCACCATGTGC 703

Db 18337 CTGATCCGCTCGCTCGGCTCCCAAGTGTGGGATTTACAGTGTGAGCCACCATGCT 18396
Qy 704 GGCCTACTTATTTCTTTTACATTCATCTTTCCCAATAGAATGTAAGATCCACAGGA 763
Db 18397 GGCACCAACAGCGCTTTTATACAGGTGGCTCATTTACCTGATTTCTCTCTCCAAATTC 18456
Qy 764 TTACTGCTATTTTCTTCTTCTTTTGTGAGACAGAGTCTCAGTTCATCAGCTCA---A 820
Db 18457 TTGTTACCCAAACACAGGAACATGAGATGGAGTCTCGCTCGCTCGCCAGTCTGGAGTGC 18516
Qy 821 COTCGTTCAGCTCACTGCAACCTCTGCTCCGCGTTCAAGYGATTTCTCTGCTCAAGC 880
Db 18517 AGTGAATCTCAGCTCACTGCAACCTCTGCTCCGCTCGCTCGCTCGCTCGCTCAGC 18576
Qy 881 CTCCTGAGTGTGGAATTTACAAAGGTGCACCAACATGCTTGGCTAAATTTTGTATTTT 940
Db 18577 CTCCCAAGTAGCTGGAGCTACAGGATGGCGGATGCCACCCACCTGGCTAA-TTTTGTATTTT 18635
Qy 941 TAGCAGAGATGGGTTTTTACCATGTTGCCAGGCTGGTCTCAAACTCCTGACCTCAAGTG 1000
Db 18636 TAGTAGAGACGGGTTTTTACCATGCTGGCCAGGCTGGTCTCGAATCTCTGACCTC--GTG 18693
Qy 1001 ATCTGCTCGCTCAGTCTCCCAAGTGTGGAATTTATAGGCGTGAGTCACTGTGCTGGC 1060
Db 18694 ATCCACCGCTCAGCTTCCCAAGTGTGGGATTTACAGGATGAGCCACCATACCTGGC 18753
Qy 1061 CGATTACTGTCTATTTT-----CTTTATGCTATATCCCGATCTAGACAGTGTCTG 1114
Db 18754 CAGGACTAGACAGTTTGTAGACATCTTTCTGAAATAGGGTGGTGGCCCTCATATTTC 18813
Qy 1115 ACATATAGTAGTGTCTCAATAAATTAATGATGAATGCACAGCTAGATATAAACTTTCTT 1174
Db 18814 ATTAATGCTATTTTCTCCAAACCGGGAACCAATAGATGTGATAGCATGCACTCCCTC 18873
Qy 1175 TTTCTTTTAAACAATCTTGACAACTTTGCGAATAAATAACAATCTGCAATCTGCT 1234
Db 18874 ATGATCTGAAGTTACTGCTCGTAATCGCGGTCCCAATTCGATTTGGATATTGGGGATGCC 18933
Qy 1235 TTTTCACATTAACA-CTTGTGTTATGACTTTTTCATATTGCTCCCAACCTTTATTTGTTACTG 1293
Db 18934 TGCAGTGTCTGTGTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 18993
Qy 1294 TTTTTCATTTGTTACTATTTTAGTCACTGAATAATATGGCTTAATTTGCTTATACATCCT 1353
Db 18994 GGGCTGGCTGGTTTTTCTCATCTTTGTTATATCTCAGCACATTTGTACATACCTGGC 19053
Qy 1354 CTGCTCCACTTTTGAAGGCCAAATTTACAATCTGATGAAGCTATGAACCTCTCCCC 1413
Db 19054 ACTCAATGAATTAATAATAATCTGAAGCGGGGTGGGGGAGCTTGTCAATATAGCCCC 19113
Qy 1414 AGAGAAATACACACACACACACTCACACAGTTTTTTTTTAATGTTTGAACATAA 1473
Db 19114 ACTTGTGTTGCTTGGCTTGAATGAAGCGGTTGTTGCTGATGATATGTCATCTCTT 19173
Qy 1474 GACAAGAACTGCAATAGAGATGTTGTTTCAATTAATTAATAAATAACTCA----- 1526
Db 19174 TTCTAGACAAATGAGTACTTGGATGAATAGGAGATCTCTTGGTTAAACCAATAAGACAG 19233
Qy 1527 GTTGGGCACTGAGTCAAGCTCTGAACACAGTACTTTTGAAGTCCAAAGTGGGTGGAT 1586
Db 19234 GCTGGCGCGTGGCTCATCTCTGTAATCCAGCACTTTGAAGCGCGAGCGGCGAGAT 19293
Qy 1587 CACTTGAAGTGAAGTTCGAGACCGCTGGTCAATATGTTGAACCGCTATCTCTACTA 1646
Db 19294 CACTTGAAGTGAAGTTCGATACCGAGCTGGCCCAACATGTTGTAACCCCAATCTCTACTA 19353
Qy 1647 AAAATACAAAATTTAGCTGGGTGATGATGATGCTGCTGATGCCAGTACTCGGAGG 1706
Db 19354 AAAATACAAAATTTAGCGAGTGGGTGGCGCATGCCCTGTTATCTAGTACTTTGGAGG 19413
Qy 1707 CTGAGGCAAGAAATGCTTGAACCTGGGAGGCGAGAGTTGCACTGAGCGGAGATCCAC 1766

Db 19414 CTGAGGAGGAGAAATCGCTTGAACCCGGGAGGTGAAGCTTTCTAGTGAGCCGAGATTGTGC 19473
Qy 1767 CACTGCACTCCAGCTCGGGGACACAGCGAGACTCTATCTCABAAAATAAATAAATAA 1826
Db 19474 CATTTACTCCAGCTCGGGGACACAGCGAGACTGTGTCTCAAAATGAATAACCCATAAG 19533
Qy 1827 ATA 1829
Db 19534 ACA 19536
RESULT 5
ACN43914/c
ID ACN43914 standard; DNA; 160482 BP.
XX ACN43914;
AC ACN43914;
XX 18-NOV-2004 (first entry)
DT Human genomic sequence HCG26773.
DE Cytostatic; carcinoma; lymphoma; cancer; human; gene; ss.
KW Cytostatic; carcinoma; lymphoma; cancer; human; gene; ss.
XX Homo sapiens.
OS WO2003073826-A2.
PN 12-SEP-2003.
PD 28-FEB-2003; 2003WO-US006235.
PF 01-MAR-2002; 2002US-00087192.
PR (SAGR-) SAGRES DISCOVERY.
PA Morris DW;
PI WPI; 2003-328604/31.
DR Recombinant nucleic acid useful for diagnosis and treatment of carcinoma
XX comprises a nucleotide sequence.
PS Claim 1; SEQ ID NO 100; Opp; English.
XX The present invention relates to novel DNA and protein sequences which
CC are associated with carcinomas. The sequences are useful for: (i) for
CC screening drug candidates; (ii) for screening of bioactive agent capable
CC of binding to Carcinoma Associated Protein (CAP); (iii) for screening of
CC a bioactive agent capable of modulating the activity of CAP; (iv) for
CC evaluating the effect of a candidate carcinoma drug; (v) for diagnosing
CC carcinoma; (vi) for inhibiting the activity of CAP; (vii) for treating
CC carcinoma; (viii) for neutralizing the effect of CAP; (ix) as a biochip;
CC (x) for diagnosing carcinoma or a propensity to carcinoma; and (xi) for
CC determining Carcinoma Associated (CA) gene copy number. In addition, the
CC CA genes are useful as DNA vaccines and the CAP are useful as markers of
CC carcinoma including lymphoma. The present sequence is one such CA coding
CC sequence. Note: This patent is an equivalent to basic patent
CC US2002182586A1, for which no sequence data was published
XX
SQ Sequence 160482 BP; 44060 A; 32143 C; 33530 G; 49875 T; 0 U; 874 Other;
Query Match 21.7%; Score 434.2; DB 11; Length 160482;
Best Local Similarity 60.2%; Pred. No. 6e-82;
Matches 888; Conservative 1; Mismatches 525; Indels 61; Gaps 8;
Qy 406 TTTTGTGTTGTTGTTTGTAGACAGAGGTCTTGTCTGTCTCACCAGGCGATGAGCACAGTGG 465
Db 5181 TGTGTTTTTTTTTCTCTCGATACGGAGTCTTGTCTGTGTTGCCAGGCTGGAGTGCAGTGG 5122
Qy 466 TGCACCATAGTGTACTGTGAGCTCAACCTCTCTGAGTCAAGGGATCTGTGACCTCAGC 525
Db 5121 CATGATCTCGGCTCACTGAAACCTCTGCTCTCTGTTTCAAGTATCTCTCTGCTCAGA 5062

Qy	526	CTCCCAAGTAGCTGGGACTACGAGCGTGCACCAACGCGCTGG--CTAATTAATAAAATTT	583
Db	5061	CTCCCAAGTAGCTGGGATTACAGCGGTGAGCCACCATGCTGGCCCTAATGTTTGTGTAAT	5002
Qy	584	TTTTTGTAGAGACTGGGTCTTACTACGTTGGCCAGGCTTGTCTTAAACCTCCTGGCTCAA	643
Db	5001	TTTTTAGTAGAGAC-AGGTTTCGCTACATTGGTCAGGCTAGTCTCAAACCTCCTGACCTCAA	4943
Qy	644	GCAATCCTCTACCTTGGCATCCAAAGTCTCGGGATTACAGGGGTGAGCCACCATGTGC	703
Db	4942	GTGATCCGCTCTGCTTGATGTACAAAGTCTGGGATTACAGATGTGAGCCACCCGCGCC	4883
Qy	704	GGCTACTTATTTCTTTACATTCATCTTTCCAAATAGATGTAGATCCACAGAACAGGA	763
Db	4882	GGCCAGTCATTCACCTTCTCTTTTCTTTTCTTTTGTGAGACAGAGTCTCATTTCTGTC	4823
Qy	764	TTACTGCTATTTTCTCTCTTTCTTTTGTGAGACAGAGTCTCACITTCATCACCTCAACCT	823
Db	4822	ATCCAGGCTGGAGTGT-----AGTGGCAC	4799
Qy	824	CCGTTACAGCTCACGTCAACCTCTGCTCCCGGGTTCAAGYGATTTCTCCTGCCCTAAGCCTC	883
Db	4798	GATCTCGGCTCACTTCAACTTTCACTCCCGGGTTCAAGTGATTTCTCCTGCCCTCAGCCTC	4739
Qy	884	CTGAGTAGCTGGAAATTACAGCGTGCACCAACATGCTTGGCTAAATTTTTTGTATTTTTAG	943
Db	4738	TTGAGTAGCTGGGATTACAGCATGCACCAACCGCACCCGGCTAAATTTTTTATATTTTTTGG	4679
Qy	944	CAGAGATGGGGTTTTACATGTTGCCCAGAGCTGTTCTCAAACCTCTGACCTCAAGTGATC	1003
Db	4678	GAGAGATGGGGTTTTCAACATGTTAGACAGGCTGCTCTCAAACCTCCTGACCTCAAGCAATC	4619
Qy	1004	TGCTGCTCAGTCTCCCAAAGTCTCGAATTATAGCGGTGAGTCACTGTGCGCTGCGCCGA	1063
Db	4618	TGCTCGGCTTGGCCTCTCTAAAGTCTGGGATTACAGCGGTGAGCCACACACCGGCGCAA	4559
Qy	1064	TTACTGTCTATTTTCTTTTATTTGCTATATCCCCAGATCTAGACGAGTGTGTGACATATAGT	1123
Db	4558	--ATCATTTCACTTTTCTACAACAGTTCGTCTGTATGCCAATGAATCTCAGGACTGTACT	4501
Qy	1124	AGGTGCTCAATAAATAATTGATGAATGACAGCGCTAGATATAAACTTCTCTTTTCTTTTT	1183
Db	4500	TCTAGCTCTAACCTCTAAACATATATTTAAGTGTCTTTAAAGCCAGGACTTAGATTCTCT	4441
Qy	1184	TTAAAACAATCTTGACAACTTTGAGAAATAATACAACTTTGCAATCTGCTCTTTTCACTT	1243
Db	4440	CAGAACTCTCAAAATTCACAAATCCAAATTCACATATATTTCCCTTAAATTTCTGCCCT	4381
Qy	1244	ATCACCTTGTTA-----TGACTTTTTTATATTCCTCAAACTTTTATTTGTTACTGTTTT	1297
Db	4380	TTCTCTGTAATTTCCCTAGTTTGTCTTAATAGTATCACTTAATTCAGGAAGAAATTTCAA	4321
Qy	1298	TTCAATTTGTTACTATTTTATGTCACCTGAAT--AATATGGCTTAATTTGCTTATACATCTCC	1355
Db	4320	GTCACTCTTCAATTTTCTCTCTCAAGCAATAAAGCACTAAATTCCTGAAGAGCTCAACC	4261
Qy	1356	TGCTCCACTTTTAGAAGGCCAAATTTACAAATCTGATGAAGCTATGAACCTCTCTCCCGAG	1415
Db	4260	TTAAACAGTCTCTCGAATGAGATGAAGGTTAACAGTTACTGAACAACAGACCAAACTAG	4201
Qy	1416	AGAAATACACACACACACACTCACACAGTTTTTTTTTAAATGTTTGCACCTAAGA	1475
Db	4200	GCACAGCAGCAAGGTTGTTCAATTGAATTCATCTCTCAGAAGCCCTAAGGAATAGCT	4141
Qy	1476	CAAGAAAACCTGANTTAGAGGATGTTTGTTCATATTAATTAATAAAT-----AATCTAGTT	1529
Db	4140	CCAATTTTATAGAGGAAGAACTGAGGTTTCCAAAGATGAAGAAATTCACACAGGAGGCT	4081
Qy	1530	GGGCACAGTACTCAAGCTGTAAACACAGTACTTTTGAAGTCCAAAGTGGGTGGATCAC	1589
Db	4080	GGGTGTGTGGCTCAGCTGTAACTCTAGCACCTTTGGAGGGCCAAAGGGGGAGTCTAC	4021
Qy	1590	TTGAGGTGAGAAGTTTCGAGACGAGCCTGGTCAATATGTTGAAACCTCTACTACTAAAA	1649

QY	1266	ATATATGCGCTCAAAACCTTTATTTGTTACTGTTGTTTTTTTTCATTGTTACTATTTTATTTAGTCACTGAAT	1325
Db	34433	CTATATTGCCCAGGCTGGTCTCAAACCTCTGGCCTCAAGTGATCTCCCTTCTGGGCTTCA	34374
QY	1326	AATATGGGCTTAATTTGCTTTATACATCCCTCTGCTCCACCTTTAGAAGGCCAAATTTACAAA	1385
Db	34373	AAAAGTGGCTGGGATTTAAGTATGAGCCACACCTCCGGGCTTCCCTCTTCCCTTCCCTTTC	34314
QY	1386	TCGTGATGAAGCTATGAACCTCTCTCCAGAGAAATACACACACACACACACTCACAC	1445
Db	34313	CTTTCTGCGCTTCCCTCTCTCTTTCTCAAAAAAATAAAGAAAGTCAGTGTGTCTGA	34254
QY	1446	ACAGTTTTTTTTTAAATGTTTGCACCTAAGACAAGAAACCTGCATTTAGAGGATGTTTGTTC	1505
Db	34253	CTGGTATTGCTCTGATTACCTGCATATATTGATATATTTCCAATAAAGATGCTTGTCTTT	34194
QY	1506	ATA--TTAATTAATAAATACTCAGTTGGGCACAGTACTCAAGCTGTAAACCAGATACT	1563
Db	34193	TTCTTTTGGTTAAAAAGTATCCGGCCAGGCGCGATGGCTCAGCTTGTAAACCCAGTACT	34134
QY	1564	TTGGAAGTCCAAAGTGGGTGGATCACTTTGAGGTGAGAAAGTTTCGAGCCAGGCTTGTCAT	1623
Db	34133	TTGGGAGGCTGTGGTGGCGGATCACCTGAGATCAGGATCAGGAGTACGAGACAGCTGGCCAA	34074
QY	1624	ATGGTGAACCCCTATCTCTACTTAAATAACAAAAATTAGCTGGGTGTAGTGATGATGCC	1683
Db	34073	ATGGTAAACCCCATCTCTACTAGAAATACAAAAATTAGCTGAGTGTAGTGCGACGCACC	34014
QY	1684	TGTAGTCCCAAGCTACTCCGGGCGCTGAGGCAAGAGAAATTTGTTGAACCTGGGAGGCGAG	1743
Db	34013	TGTAATCCCAAGCTACTTTGGGAGGCCAAGGAGAGAAATTTGTTGAACCCGGGAGGCGAG	33954
QY	1744	GTTCGAGTGAGCGGAGATCCCACTGCATCTCCAGCTGGGCGGACAGCGAGACTCTA	1803
Db	33953	GTTCGAGTGAGCGGAGATCACACCAATTGCACTCCAGCCAGGCGGACA-AGAGAGACTCTG	33895
QY	1804	TCTCAAAAAATAAATAAATAAT	1828
Db	33894	TCTCAAAAAATAAATAAATAAAT	33870
RESULT 8			
ID	ABK83569	standard; cDNA; 122888 BP.	
XX	AC	ABK83569;	
XX	AC		
DT	14-AUG-2002	(first entry)	
XX	DE	Human cDNA differentially expressed in granulocytic cells #140.	
XX	KW	Human; ss; granulocytic cell; DNA chip; bacterial infection;	
KW	KW	viral infection; parasitic infection; protozoal infection;	
KW	KW	fungal infection; sterile inflammatory disease; psoriasis;	
KW	KW	rheumatoid arthritis; glomerulonephritis; asthma; thrombosis;	
KW	KW	cardiac reperfusion injury; renal reperfusion injury; ARDS;	
KW	KW	adult respiratory distress syndrome; inflammatory bowel disease;	
KW	KW	Crohn's disease; ulcerative colitis; periodontal disease;	
XX	KW	granulocyte activation; chronic inflammation; allergy.	
OS	XX	Homo sapiens.	
XX	XX	WO200228999-A2.	
PN	XX		
XX	PD	11-APR-2002.	
XX	XX		
PF	XX	03-OCT-2001; 2001WO-US030821.	
XX	XX		
PR	XX	03-OCT-2000; 2000US-0237189P.	
XX	XX		
PA	XX	(GENE-) GENE LOGIC INC.	
PI	XX	Beazer-Barclay Y, Weissman SM, Yamaga S, Vockley J;	

QY 1291 CTGTTTTTTCATGTTA CTATTTTAGTCACTCAATAAATATAGGCTTAATTTGCTTATACAT 1350
 Db 98162 ATATAATAACCAAGGTTCTTATTTATGTTATCCATTGCTTGAATTAACAATGTTTTCATC- 98104
 QY 1351 CCTCTGCTCCACTTTAGAGGCCAAATTTTCAAAATCTGATCAAAAGCTATGACCTCTC 1410
 Db 98103 -----AAACTATAAAAGACCTTTAAACAAGGTAAGGATGAGAGGAACCTTGAAGTTGC 98051
 QY 1411 CCCAGAGAATAACACACACACACACACTCACACACAGTTTTTTTTTAAATGTTTGCAC 1470
 Db 98050 CCAAGATAAAATTTTGTCTCAAGTCTCTCACTTTGTTTCATGTTTAAATCTCAATA 97991
 QY 1471 TAAGACAGAAACCTGCAATTAGAGGATGTTTTCATATTAATTAATAAATAA----CTCA 1526
 Db 97990 TATTTAATCTATTGCAATTA CAATAATTTTGTGTCACAAAAAATAAGCTCCTGG 97931
 QY 1527 GTTGGGCACAGTCACTCAAGCTGTAACACACAGTACTTTTGAAGTCCAAAGTGGTGCAT 1586
 Db 97930 GCTGGACACGGTGGCTCAGCCTGTAATCCAGCACTTTGGAGCCGAGGAGGTTGGAT 97871
 QY 1587 CACTTGAGGTGAGAAAGTTTCAGACCAAGCTGCTCAATATGTTGAAACCTTATCTCTACTA 1646
 Db 97870 CAC--CAGGTCAGGAGATGGAGACCATCTTGCCCAACATGTTAAACTCTGCTCTACTA 97813
 QY 1647 AAAATACAAAATTAGCTGGGTGATGATGATGATGATGATGATGATGATGATGATGATGATG 1706
 Db 97812 CAAATACAAAATTAGCTGGGTGATGATGATGATGATGATGATGATGATGATGATGATGATG 97753
 QY 1707 CTGAGGCAAGAAATTTGCTTGAACCTGGAGGACAGAGTTGCGAGTCCAGATCCAC 1766
 Db 97752 CTGAGGCAAGAAATTTGCTTGAACCTGGAGGACAGAGTTGCGAGTCCAGATCCAC 97693
 QY 1767 CACTGCACTCCAGCTGGGCGACACAGGAGACTCTATCTCAAAAAATAA 1817
 Db 97692 CACTGCACTCCAGCTGGGCGACAGAGAGACTCTATCTCAAAAAATAA 97642

RESULT 12
 ID ADP68568/c
 XX ADP68568 standard; cDNA; 86000 BP.
 AC ADP68568;
 XX
 DT 09-SEP-2004 (first entry)
 XX
 DE Human PPAR-alpha cDNA.
 KW cytostatic; gene therapy; PPAR-alpha;
 KW peroxisome proliferator-activated receptor-alpha; PPAR-alpha modulator;
 KW PPAR-alpha associated disorder; hyperproliferative disorder; human; gene;
 KW ss.
 XX
 OS Homo sapiens.
 XX
 FN US2004115637-A1.
 XX
 PD 17-JUN-2004.
 XX
 PF 11-DEC-2002; 2002US-00317500.
 XX
 PR 11-DEC-2002; 2002US-00317500.
 XX
 PA (ISIS-) ISIS PHARM INC.
 XX
 PI McKay R, Dobie KW;
 XX
 DR WPI; 2004-449378/42.
 XX
 PT New oligonucleotide compound that inhibits expression of PPAR-alpha,
 PT useful for preparing a composition for treating hyperproliferative
 PT disorders, e.g. cancer.
 XX

PS Claim 1; SEQ ID NO 4; 121pp; English.
 XX
 CC The invention describes a compound, having a sequence comprising 8-80 bp
 CC targeted to a nucleic acid encoding PPAR-alpha (peroxisome proliferator-
 CC activated receptor-alpha), that specifically hybridizes with the nucleic
 CC acid encoding PPAR-alpha, comprising 86001-bp sequence and inhibits
 CC expression of PPAR-alpha. Also described are: a method of inhibiting the
 CC expression of PPAR-alpha in cells or tissues; a method of screening for a
 CC modulator of PPAR-alpha; a diagnostic method for identifying a disease
 CC state; a kit or assay device comprising the compound; and a method of
 CC treating an animal having a disease or condition associated with PPAR-
 CC alpha. The oligonucleotide compound is useful for preparing a composition
 CC for treating hyperproliferative disorder e.g. cancer. This sequence
 CC represents a human peroxisome proliferator-activated receptor-alpha (PPAR
 CC -alpha) cDNA.
 XX
 SQ Sequence 86000 BP; 22822 A; 19623 C; 20377 G; 23178 T; 0 U; 0 Other;
 Query Match 20.7%; Score 413.6; DB 12; Length 86000;
 Best Local Similarity 61.0%; Pred. No. 1.2e-77;
 Matches 878; Conservative 1; Mismatches 490; Indels 71; Gaps 10;
 QY 405 TTTTGTGTTGTTTGTAGACAGAGGCTTGTCTCTGTCAACGATCTGTCGACCTCAG 464
 Db 69228 TTTTGTGTTGTTTGTAGACAGAGGCTTGTCTCTGTCAACGATCTGTCGACCTCAG 69169
 QY 465 GTGCAACATAGGTCACCTGACGCTCAACCTCTGAGCTCAAGGATCTGTCGACCTCAG 524
 Db 69168 GTGCAATCTAGCTCAGCTCAGCTCAACCTCTGAGCTCAAGGATCTGTCGACCTCAG 69109
 QY 525 CCTCCCAAGTAGCTGGAGTACGAGCTGACCAACGCTGCTCAATTAATAAATAATTT 584
 Db 69108 CCTCCCAAGTAGCTGGAGTACGAGCTGACCTCAACGATCTGTCGACCTCAG 69051
 QY 585 TTTTGTAGACACTGGGTTCTTACGTTGGCAGGCTTGTCTTAACTCTGGCTTCAAG 644
 Db 69050 TTTTATTAGACAGGGGTTTGTGCTATGTTGCGAGGCTGCTCAAACTCTGACCTCAAG 68991
 QY 645 CAATCTCTCTACCTTGGCATCCCAAGTGTGGGATTTACAGGGGTGAGCCACCATGTCG 704
 Db 68990 TGATCCACCGCTCGGCTTCCAAAGTGTGGATTTACAGGTTGTAACCAACCGCATCTG 68932
 QY 705 GCTACTTATTTTATCATTTCTTCTTCCAAATAGAAATGTAAGATCCACAGAACAGGAT 764
 Db 68931 GGAGACTTTTATTTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTT 68872
 QY 765 TACTGCTATTTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCT 824
 Db 68871 CTCCTCTGTGCGCCAGGCTGGAGTGCAGTGGCGCAATC----- 68833
 QY 825 CGTTGAGCTCAGCTGCAACCTCTGCTCCGCGGTTCAAGYATTTCTCTGCTCAAGCTCC 884
 Db 68832 ----TCAGCTCAGCTGCAACCTCTGCTCCGCGGTTCAAGYATTTCTCTGCTCAGCTCC 68776
 QY 885 TGAGTAGCTGGAATTAACAAGCTGCAACCATGCTTGGCTTAATTTTCTTCTTCTTCTTCTT 944
 Db 68775 CGAGTAGCTGGGACTACAGGCTATGGCCACCAACCCAGCTAA-TTTTTCATCTTCTAGT 68717
 QY 945 AGAGATGGGTTTATACATGTTGCCAGGCTGTCTCAAACTCTGACCTCAAGTATCT 1004
 Db 68716 AGAGAGGGGTTTACCATATTTGGCCAGGCTGTCTCAAACTCTGACCTGAGTATCTC 68657
 QY 1005 GCCTGCTCAGTCTCCCAAGTGTCTGAATTAAGGCTGAGTCACTGTGCTTGGCGCAT 1064
 Db 68656 ACCCAGCTCAGCTCCCAAGTGTCTGGATTAAGGCTGAGTCACTGTGCTTGGCGCAT 68598
 QY 1065 TACTGTCTATTTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTT 1124
 Db 68597 ----CTTTTATTTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTT 68542
 QY 1125 GGTGCTCAATAAATAATGATGAATGACAGCTAGATATAAATTTCTTTTCTTTTCTTTT 1184
 Db 68541 GCATTTATCATAGCTTGTATCTGTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCT 68489

Db	4111	AAATGGGCTTGCTCTCAAGCCACGACGCTGGTAGTGGCAGAGTTGGGAATCAAAACACAAT	4052
Qy	1250	TTGTTATGACTTTTTTCATATTTGCTCAAACTTTATTTGTTACTGTTTTTCATTTGTTACT	1309
Db	4051	TTGTTTGTATCCCAAAATCTTTGGCCCTTTCCACAAACAGGCGCTACATGTCCCTTACA	3992
Qy	1310	ATTTTGTACTGTAATATGCTTAATTTGTTGTTATATCATCTCTCTGCTCCACTTTAGA	1369
Db	3991	TATTTTTCAGGTGTACATTTATCTTTTCAGCATTTGTAAC-TCTAGTGAAGCCTTAGATC	3933
Qy	1370	AGGCCAAATTTACAAATCTGATGAAGCTATGAACCTCTCCCCAGAGAAATACACAC	1429
Db	3932	TCACATAATAGCAATGACCTCTTCTAGCTGTATTTATTTAGTGTCTTTATGATGCGCAGGC	3873
Qy	1430	ACACACACACTCCACACACAGTTTTTTTTTAAATGTTTGCACATGATCTCAATTTTACACATCCCATGAGAT	1489
Db	3872	ATTGCAATTTAGCACTTCACACATATGATCTCAATTTTACACATCCCATGAGAT	3813
Qy	1490	TAGAGGATGTTTGTTCATATTAATTAATAAATACTCAGTTGGGCACAGTCAAGCCT	1549
Db	3812	TGGCACTATTTATGCGCTTTAACAGAAAGGAAACAGGCCAGCATGGTGGCTCATGCT	3753
Qy	1550	GTAACACACTACTTTGGAGTCCAAGTGGGTGGATCACTTGAGTGAAGTTTCGAGA	1609
Db	3752	ATAATCTCACTACTTTGGGGAGGTGAGGTGGGTGATCACTCCTGAGGTGAGGATTTCAAGA	3693
Qy	1610	CCAGCTGGTCAATATGTTGMAACCTTCTCTCTACTTAAATAATACAAAATTAGCTGGGTG	1669
Db	3692	CCAACTGGCCAACTAGTGAACCCCACTCTCTACTTAAATAATACAAAATTAGCGCGCA	3633
Qy	1670	TAGTGATGATGCTGTAGTCCAGCTACTTCGGAGGCTGAGGCAAGAGAAATTTGCTTCAA	1729
Db	3632	TGTTGGCGGTGCTGTATCCAGCTACTTCGGAGGCTGACACAGGAGATTTGCTTGAA	3573
Qy	1730	CTTGGAGGACAGGTTGACGTGAGCCGAGATCCACACTGCTCACTCCAGCTGGCGAC	1789
Db	3572	CCAGGAGGCGAGGTTGCACTGAGCAGAGATTTGCACTTTCGACTCCAGCTGGAGAC	3513
Qy	1790	ACAGCGAGACTCTATCTCAAAAAATAAATAAATAAATAA 1831	
Db	3512	AGAGCAAGACTTCGCTCTCAAAAGAAACAGAGAGGAAACAGA 3471	

RESULT 14

AAAL42369/c

ID AAL42369 standard; DNA; 13670 BP.

AC AAL42369;

XX 28-JUN-2002 (first entry)

XX Human Guanine nucleotide binding protein gamma 7 (GNG7) gene sequence.

XX Human; gene; db; guanine nucleotide binding protein gamma 7; G protein;

XX GNG7; novel polymorphic site; drug screening; gene therapy;

XX GNG7-related disease; pancreatic cancer; GNG7 haplotyping;

XX GNG7 genotyping.

XX Homo sapiens.

XX Key Location/Qualifiers

FT replace(3900, T)

FT /*tag= a

FT /notes= "Single nucleotide polymorphism"

FT replace(3975, G)

FT /*tag= b

FT /notes= "Single nucleotide polymorphism"

FT 4005..9671

FT /*tag= c

FT /product= "Guanine nucleotide binding protein gamma 7"

FT 4005..4085

FT /*tag= d

FT exon

FT intron

FT /number= 1

FT 4086..9545

FT /*tag= e

FT /number= 1

FT 9546..9671

FT /*tag= f

FT /number= 2

FT replace(9958, A)

FT /*tag= g

FT /note= "Single nucleotide polymorphism"

XX W0200218647-A1.

XX 07-MAR-2002.

XX 23-AUG-2001; 2001WO-US026279.

XX 25-AUG-2000; 2000US-0228234P.

XX (GENA-) GENAISSANCE PHARM INC.

XX Finkel K, Kliehm SE, Koshy B;

XX WPI; 2002-315548/35.

XX P-PSDB; AA014753.

XX Novel genetic variants of guanine nucleotide binding protein, gamma 7

XX gene useful in studying expression and function of the protein, and for

XX screening drugs to treat diseases e.g. pancreatic cancer.

XX Claim 1; Fig 1; 62pp; English.

XX The invention comprises the nucleotide and amino acid sequences of the

XX human guanine nucleotide binding protein (G protein) gamma 7 (GNG7). The

XX invention specifically relates to the discovery of three novel

XX polymorphic sites in the GNG7 gene. The GNG7 nucleotide and protein

XX sequences are useful for screening for drugs which target GNG7 and may

XX be used to treat (gene therapy) GNG7-related diseases (e.g. pancreatic

XX cancer). The GNG7 nucleotide sequence can be used for haplotyping GNG7,

XX genotyping GNG7 and predicting the haplotype pair for GNG7 in an

XX individual. The present DNA sequence represents the human GNG7 gene

XX sequence

SQ Sequence 13670 BP; 3047 A; 3946 C; 3680 G; 2997 T; 0 U; 0 Other;

Query Match 20.7%; Score 413.4; DB 6; Length 13670;

Best Local Similarity 60.6%; Pred. NO. 8.9e-78;

Matches 876; Conservative 1; Mismatches 512; Indels 56; Gaps 10;

Qy 404 TTTTGTGTTGTTGTTTGTAGAGACAGGGTCTGTCTGTCTCACCAGGAT-GAGCAG 462

Db 5900 TTTGTTTTGTTTTTTTTTCTGAGATGGAGTCTCACCTGTGCTGGGTTGGAGGGCAG 5841

Qy 463 TGGTCAACACCTAGTCTACTGAGCCTCAACCTCTGAGCTCAAGGATCTGTGACCTC 522

Db 5840 TGGCGCAATCTCAGCTCACTGCAACCTCTGCTCTCGGGTTCAAGTATTCTCTGACTC 5781

Qy 523 AGCTCTCCCAAGTAGCTGGGACTACGAGGCTGCACACAGCCTGGCTAATTAATAAAT 582

Db 5780 AGTCTCCCAAGTAGCTGGGACTATAGGAGCCTGCGCACCAAGCTGGCTAATTTTGTGA 5721

Qy 583 TTTTGTGTTAGAGCTGGGTCTTACTAGTTGGCCAGAGCTGTCTTAACCTCTGGCTTCA 642

Db 5720 ATTTTAGTAGAGATGGGGTTTCCACCATATATGGCAGGCTGTGCTCGAACCTCTGAC--CT 5663

Qy 643 AGCAATCTCTCTACCTTGGCATCCCAAGTCTGGGATTCAGGGGTGAGCCACCATGTG 702

Db 5662 TGTGATCCGCCCACTCAGCTCCCAAGTCTGCGATTCAGGCGTGAGTCAACCGTCC 5603

Qy 703 CGGCTACTTATTTCTTTTACATTTCCATCTTTTCCAAATAGAAATGATGATCCACAGACAGG 762

Db 5602 TGGCCAGTTAAATGTTTTTTTTTAGATGTTGCTGTGAAGGTTGTTTTTTC----- 5554

Qy	622	TGTCCTTAAACTCTCTGGCTTCAAGCAATCTCTCTTCACTTGGCAATCCCAAAAGTCTCGGATT	681
Db	4745	GGTCTCCATCTCTCTGACCTC--GTGATCCACTCGCCTTCGGCTTCCCAAAAGTCTCGGATT	4802
Qy	682	ACAGGGGTGAGCCACCATGTGCGGCTACTTATTTCTTTTACATTTCCATCTTTTCCAATAGAA	741
Db	4803	ATGGGGGTGAGCCACCGCACCCGGCAGGGGTTTTTTCTGTTTTTTTTTTTTTTTTTTTTTTTT	4862
Qy	742	TGTAAAGATCCACAGAACAGGGATTACTTGCCTATTTTTCTTCTTTCTTTTGTGAGACAGAG	801
Db	4863	TTTTTTTTTTTTTTTTTTTTTTTTTAAAGACAGACTCTCGCTCTGTCAACCGCTAGAGTGCAG	4922
Qy	802	TCTCACTTCATCACTCA--ACCTCCGTTCACTCTCACTCGAACTCTCGCTCCCGGGTTCA	860
Db	4923	TGACACAACTCTTGGCTCACTGCAACCTCCAACTCCCTGCAAACTCTGCTCCCGGGTTCA	4982
Qy	861	AGVGAATCTCTGGCTTAAGCCTCCTGAGTAGCTGGAATTACAGCGTGCACCAACCATGCT	920
Db	4983	AGCAACCTCTCTGCTCAGCCTCTGAGTAGCTGGAATTACAGCGCCTTGCCACTGCACC	5042
Qy	921	TGGCTAATTTTTTGTATTTTTTAGCAGAGATGGGGTTTTTACCATTTTGCCTCAGCGTCTCT	980
Db	5043	CAGCTAA--TTTTTGTATTTTTTAGTAGAGCGGGTTTCATCAITGTTGGCTAGGCTAGTCT	5101
Qy	981	CAAACTCCTGACCTCAAGTGATTCGCTCGCTCAGTCTCCCAAAGTCTGGAATTTATAGG	1040
Db	5102	CAAACTCCTGACCTCAGATGATCCATTGGCCTTGGTGTCTCAAAAGTCTGGGATTACAGG	5161
Qy	1041	CGTGAGTCACTGTGCTGGCGGATTACTGTCTATTTTTCTTTATTTGCTATATCCCCAGATC	1100
Db	5162	CATGAGCCACCACTCAGCCTGGT--TTAAGCATTTTTTATATGACTCAACTCATTTAACCC	5220
Qy	1101	TAGAGCAGTGTCTCACATATAGTAGTGCTCAATAAATAATTGATGAATGCACAGCCTAG	1160
Db	5221	GCCCCAACTCCCTCAGGATGTCTGTGTAAATGATGCCCACTTACAGTTAAATTAAG	5280
Qy	1161	ATATAAACTTTCTTTTTTTTAAACAACTCTGCACAACTCTGCAGAAATTAATAACAA	1220
Db	5281	CCCAGAAAG-----GTTAAGGGACCTGCCTGAGGCTGCACAGCACTCTGAA	5328
Qy	1221	TCTTGCACTTCTGCTTTTTCACCTTATCACTTGTATGACACTTTTTTCATATTCCTCAAC	1280
Db	5329	CCCTGGCTCTCGGCAITCTCTTAGCTGCTGTGTA-----GTGGCCCAAAATC	5376
Qy	1281	TTTATTTGTTACTGTTTTTTCATTTCTTACTATTTTTTAGTCACTGTAATAATGCGCTTA	1340
Db	5377	AAATGTTATTAATTTAGAAATTCAGGTTATTTGTTGAAATAGCCCAATTTAGCTTTAAA	5436
Qy	1341	GCTTATACATCCTCTGCTCCACTTTTGAAGGCCAAATTTTCAAACTGATGAAGCTAT	1400
Db	5437	TTTAAACCAATGTTGTTTAAACATTTAGAAATGTGCACCTTAAATAACACCAATTTCTGG	5496
Qy	1401	GA--ACCTCTCCCMAGAAATACACACACACACACTCTCACACAGCTTTTTTTTTTTA	1459
Db	5497	CACAGTGGCTCACGCTCTTAATCCCAGCACTTTGGGAGGCCAAGGCGAGGACATCACTTG	5556
Qy	1460	ATGTTTTCGAATTAAGACAGAAACCTGCATTAGAGGATGTTGTTCATATTAATAAAA	1519
Db	5557	AGGTGAGGAGTTTCGAGACCAAGTCTGGCCAAACATGTTGAACCCCAATCTCTATTA	5616
Qy	1520	T-----AACTCAGTTGGGCACAGTGACTCAAGCCTGTAAACACAGTACTTTGGAGTCCAA	1575
Db	5617	TATAAAATTAGCTGGGCACGCTGGCTCACACCTGTAAATCCAGCACTTTGGGAGGCTGA	5676
Qy	1576	GGTGGGTGGATCACTTGAGGTGAGAAATTCAGACACCGCTGGTCAATATGTTGAACCC	1635
Db	5677	CACGGGTGGATCA--TGAGGTGAGGAGATCGCGACCATCTCTGGCTTAACCGTGAACCC	5734
Qy	1636	TATCTCTACTAAAAATACAAAAATTTAGCTGGGTGTAGTGATGTCCTGTAGTCCAGC	1695
Db	5735	TGTCCTTACAAAAATACAAAAATTTGGCGGGCGTGGTGGCAGGTTGATGTAGTCCAGC	5794
Qy	1696	TACTCGGGAGGCTGAGGCAAGAGAAATTCCTTTGAACCTCGGAGGCGAGGTTGCGATGAGC	1755

GenCore version 5.1.7
Copyright (c) 1993 - 2006 Bioceleration Ltd.

OM nucleic - nucleic search, using sw model

Run on: February 9, 2006, 13:34:47 ; Search time 9440.43 Seconds
(without alignments)
9917.025 Million cell updates/sec

Title: US-10-607-806-1_COPY_7200_9200

Perfect score: 1998.6

Sequence: 1 ctatctcgctgtccctccc.....ckctacataaatatttatta 2001

Scoring table: IDENTITY NUC

Gapop 10.0 , Gapext 1.0

Searched: 41078325 seqs, 23393541228 residues

Total number of hits satisfying chosen parameters: 82156650

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : EST:*

1: gb_est1:*
2: gb_est2:*
3: gb_est3:*
4: gb_est4:*
5: gb_est5:*
6: gb_est6:*
7: gb_est7:*
8: gb_est8:*
9: gb_est9:*
10: gb_est10:*
11: gb_est11:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
c 1	505.4	25.3	507	5	BU785040 in43g09.x
c 2	344.2	17.2	4087	4	BC024593 Homo sapi
c 3	341	17.1	3990	4	AL713681 Homo sapi
c 4	337.2	16.9	15970	9	AG839852 260L13-C5
c 5	315.8	15.8	897	6	CD557847 AGENCOURT
c 6	311	15.6	3095	4	CR858326 Pongo pyg
c 7	307.2	15.4	5797	4	CR749233 Homo sapi
c 8	301.2	15.1	1605	4	CR860521 Pongo pyg
c 9	301.2	15.1	1863	4	CR599842 AGENCOURT
c 10	300.6	15.0	946	5	BQ958903 AGENCOURT
c 11	298	14.9	918	5	BQ706343 AGENCOURT
c 12	293.6	14.7	736	8	CR780936 Homo sapi
c 13	290.6	14.5	3092	4	CR613629 full-leng
c 14	288.2	14.4	855	5	BQ681302 AGENCOURT
c 15	287.6	14.4	769	5	BQ710315 AGENCOURT
c 16	285.2	14.3	4088	4	CR648311 Homo sapi
c 17	283	14.2	3165	4	CR859576 Pongo pyg
c 18	281.6	14.1	800	1	AUL20942 AUL20942
c 19	281.4	14.1	3259	4	CR860168 Pongo pyg
c 20	280.4	14.0	672	6	CA431692 UI-H-DF0
c 21	280	14.0	3474	4	AL110229 Homo sapi
c 22	278.8	13.9	658	9	AQ393450 CITBI-E1

c 23	277.6	13.9	877	9	AQ739838 HS 5505 A
c 24	277.2	13.9	1875	4	BC009270 Homo sapi
c 25	276.6	13.8	3990	4	AL713681 Homo sapi
c 26	276.4	13.8	4828	4	AL390128 Homo sapi
c 27	276.2	13.8	881	5	BUS21286 AGENCOURT
c 28	274.8	13.7	3552	4	CR860263 Pongo pyg
c 29	273.8	13.7	736	6	CA427039 UI-H-DF0
c 30	273.4	13.7	629	5	CR859082 Pongo pyg
c 31	273.4	13.7	2230	4	CR859082 Pongo pyg
c 32	272.4	13.6	652	6	CA427045 UI-H-DF0
c 33	272.4	13.6	666	6	CA431783 UI-H-DF0
c 34	272.4	13.6	815	8	CX785622 HES3_49
c 35	272.2	13.6	922	5	BUS01973 AGENCOURT
c 36	272	13.6	617	3	BI861844 603388872
c 37	271.8	13.6	644	7	CN480313 UI-H-EU0
c 38	271.6	13.6	1042	5	CR377759 BX377759
c 39	271.6	13.6	2097	4	CR614786 full-leng
c 40	271.4	13.6	603	1	AL707313 DKFp686p
c 41	271.2	13.6	5325	4	HSM804049 Homo sapi
c 42	271.2	13.6	5785	4	AL137734 Homo sapi
c 43	270.6	13.5	2330	4	CR615928 full-leng
c 44	270.4	13.5	618	9	BZ609884 WHACH76TF
c 45	270	13.5	665	5	BU633001 UI-H-DF0

ALIGNMENTS

RESULT 1
BU785040/c
LOCUS in43g09.x1 HR85 islet Homo sapiens cDNA clone IMAGE:6125008 3', linear EST 11-OCT-2002
DEFINITION BU785040 mRNA sequence.
ACCESSION BU785040
VERSION BU785040.1 GI:23830576
KEYWORDS EST.
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 507)
AUTHORS Melton,D., Brown,J., Kenty,G., Permutt,A., Lee,C., Kaestner,K., Lemishka,I., Searce,M., Brestelli,J., Gradwohl,G., Clifton,S., Hillier,L., Marra,M., Pape,D., Wylie,T., Martin,J., Blistain,A., Schmitt,A., Theising,B., Ritter,S., Ronko,I., Bennett,J., Cardenas,M., Gibbons,M., McCann,K., Cole,R., Tsagarishvili,R., Williams,T., Jackson,Y. and Bowers,Y.
TITLE Endocrine Pancreas Consortium
JOURNAL Unpublished (2000)
COMMENT Contact: Douglas Melton, Klaus H. Kaestner, & Hiroshi Inoue
Endocrine Pancreas Consortium
Harvard University, Howard Hughes Medical Institute
Dept of Molecular and Cellular Biology, 7 Divinity Ave, Cambridge, MA 02138
Tel: 617-495-1812
Fax: 617-495-8557
Email: dmelton@biohpc.harvard.edu
Library was constructed by Dr. Hiroshi Inoue DNA sequencing by: Washington University Genome Sequencing Center For information on obtaining a clone please contact: Dr. Hiroshi Inoue (hinoue@im.wustl.edu)
Seq primer: -40UP from Gibco
High quality sequence stop: 443.
FEATURES
source
1. 507
Location/Qualifiers
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/clone="IMAGE:6125008"
/tissue_type="Purified pancreatic islet"
/lab_host="DH10B"
/clone_lib="HR85 islet"

REFERENCE
AUTHORS

1. (bases 1 to 4087)
Straussberg, R.L., Feingold, E.A., Grouse, L.H., Derge, J.G.,
Klausner, R.D., Collins, F.S., Wagner, L., Shenmen, C.M., Schuler, G.D.,
Altshuler, S.F., Zeeberg, B., Buetow, K.H., Schaefer, C.F., Bhat, N.K.,
Hopkins, R.P., Jordan, H., Moore, T., Max, S. I., Wang, J., Hsieh, F.,
Diachenko, L., Marusina, K., Farmer, A. A., Rubin, G.M., Hong, L.,
Stapleton, M., Soares, M.B., Bonaldo, M.F., Casavant, T.L.,
Schetz, T.M., Brownstein, M. J., Usdin, T. B., Toshiyuki, S.,

Query Match	17.2%	Score	344.2	DB 4	Length	4087
Best Local Similarity	58.3%	Pred. No.	5.5e-22	Indels	69	Gaps
Matches	859	Conservative	1	Mismatches	544	
Qy	399	TTGGGTTTTTTTTTGTGTTGTTTAGAGACAGGGTCTTGCTCTCACCCAGGCAATGAGC	458			
Db	2019	TTTGAAATTTGTTTGTGTTTGTGAGACAGGGTCTGGGTCTGCCACCCAGGCTGGAGT	2078			
Qy	459	ACAGTGTGCAACCATGAGGTCACTGCAGCCTCAACCTCCTGAGCTCAAGGATCTGCTGA	518			
Db	2079	GTAGTGTGTGATCATAGCTCAATGCGAGCCCCAACCTTCGGGGCGCAAGTGATGCTCTCA	2133			
Qy	519	CCTCAGCCTCCCAAGTAGCTGGGACTACGAGCGTGCACACACGCCTCGGCTAAATTTAAA	578			
Db	2139	CCTCAGCCTCTCTGAGTAGCTGGGACTACAGGCAATGCCACACACCCCTCTAATATATA	2198			
Qy	579	AAAT-----TTTTTTGTAGAGACTGGGCTTTTACTAGTTGGCCAGGCTTGCTTTAAACTC	633			
Db	2199	TTTTTTTTTATTTTTTGTAGACTGAGGTTTCATTATGTGTCAGACTGGTCTGGAATCTC	2258			

Qy	634	CTGGCTTCAGCAATCCTCCTACCTTGGCATCCCAAAGTGTCTGGGATTACAGGGGTGAGC	693
Db	2259	CTGGC-TCAAAGTGATCTCCTCTCTGCCCTCCCGAAGTGTGGGATTGGGTGACGAGC	2317
Qy	694	CACCATGTGGGGCTACTATTTCCTTTATCATTCCTCATCTTTCCAAATAGAAATGTATAGATCCAC	753
Db	2318	CACCACATCTGGCCCTGTTGATTATTAATTTTTTTTTTTT-----	2356
Qy	754	AGAACAGGATTAATCGCTATATTTTCTTCCTTTCTTTTGTGAGACAGAGTCTCACTTCATC	813
Db	2357	-----TTTTTTTTTGTGAGATAGAGTCTCGCTCTCGCTCAGGCTGAGT	2402
Qy	814	ACCTCAACCTCGTTTCAGCTCAGTCGAACCTCTGCCTCCCGGTTCAAGTGATCTCCTTG	873
Db	2403	GCAGTGTGTGATCTCGGCTCGCTGCAACCCCGCTCTCCCGGTTCAAGTGATTTTCCCTG	2462
Qy	874	CCTAAGCCTCTCTGATGAGTGGGAATTAACAAGCGTGACCACTGCTTGCTGCTCAATTTTTT	933
Db	2463	TGTCAGCCTCCGAGTAGCTGGGGTTACAGGTCTCACACCAAGCCTGCTGCTAA-TTTTT	2521
Qy	934	GTATTTTTAGCAGATAGGGGTTTTCAATGTGTGCCCAGGCTGGTCTCAAACTCTCTGACC	993
Db	2522	GTATTTTTTAGTACAGCTGGGGTTTTCACCATGTTGGCCAGGCTGCTCGAACTCTTGACC	2581
Qy	994	TCAAGTGATCTGCTCGCTCAGTCTCCCAAGTGTGGAATATAGCGGTGAGTCACTGT	1053
Db	2582	TC--GTATCTGCCCTCGCTCGGCTCCCAAGTGTGGAATACAGGTGTGAGGCCACCGT	2639
Qy	1054	GCCTGGCCGATTACTGTCTATTTTTTTTATTTGCTATATCCCCAGATCT-AGAGCAGTGTCT	1112
Db	2640	GCCAGCCGATTAATTAATATTTTATTTGCTTCCCTTTTGCAGAAATGATTTGATTGCT	2699
Qy	1113	TGACATATAGTAGTGTCTCAATAAATTAATGTAGTAATGACAGCCTAGA----TATAAAC	1168
Db	2700	AGGCTTATATTCAGCCATCAATAACCAAAATGTCTGAGCCCTGACTCAGGGCTTTTACAT	2759
Qy	1169	TTTCTTTTTCTTTTTTAAACAATCTTGACACTTTTGCACTTTTGCAATTAAT-----	1216
Db	2760	CTATCACTGTGGTATGATTTCTTCTTAGGAGGCTTTTGCAAGTGTGGAATAGTCTGCCTT	2819
Qy	1217	ACAACTTGTCAATCTGCTTTTTTCACTTATCACCTTGTATGACTTTTTCATATGCTCTCA	1276
Db	2820	ATAATAGTAAGACAAACTTTTATGATTTTATGCCAATATGAGTAGGAAAAAGTCCAG	2879
Qy	1277	AACCTTTATGTTACTGTTTTTTTCAATGTTACTATTTTATGTCAGTGAATATATGCTT	1336
Db	2880	AGAAATCTGTATATAATAATCTGATTTCTGTGGGTTTTTTTTGTAGGTTTCAATTTT	2939
Qy	1337	ATTTGCTTATACATCCTCGCTCCACTTTTAGAAGGCCAATTTTCAAAATCTCATGAAAG	1396
Db	2940	AGCTGGGCAATCATCTCTTCTGCCAGGGGTATGGGTCAATTTT---TTTTTCAAGAGGG	2996
Qy	1397	CTATGAACCCCTCTCCGAGAAATACACACACACACACTCTCACACAGTTTTTTTTT	1456
Db	2997	AAATTAATTTTGTGTACAAAGTAATTTGATTTGTGGTATCTATAAGAAATACTTCAA	3056
Qy	1457	TTAATGTTTGCACACTAAGACAGAAACCTGCATTAGAGGATGTTTGTTCATATTAATA	1516
Db	3057	CTGAAGAGTTTTAAAAATGTGTAGAAATACAGAACTCGAGCTTCACTCTGCGCAAGATGC	3116
Qy	1517	AAATAACTAGTTGGGCAGTCACTCAAGCCCTGTAACCAAGTACTTTTGGAGTCCAAG	1576
Db	3117	AAATTTAGGCC---GGGTGCAGTGACTACGCCCTGCAGTCTTTAGCCCTTTTGGGAGGCCGA-	3172
Qy	1577	GTGGGTGGATCACTTGAGGTGAGAAGTTGAGACCAAGCCTGGTCAATATATGGTGAACCCCT	1636
Db	3173	GTGGGAGGATGCTTGAGCTCAAGATTTTGAGACAGGGCTGAGCAATGTAGTGGACCTC	3232
Qy	1637	ATCTCTACT-AAAAATACAAAATTAGTGGGTGTAGTGTGATGTGCTGTAGTCCACAG	1695
Db	3233	GTCTCTATAAAAAAATTTAAAGGTAGCTGGTGTGGTGTGGTGTGGCCCTGTGGTCTGGC	3292

Qy	1696	TACTCGGGAGGCTGAGCGCAAGAGAAATGCTTGAACCTGGGAGCGCAGAGGTTGCAGTGAAC	1755
Db	3293	TACTTGGGAGGCTGTGATGGGAGGATCGCTGGAGCGCTGGGAGGTCAAGGCTCGGCTGGGT	3352
Qy	1756	CGAGATCCCAACCACTGCACCTCCAGAGCTGGGCGCACAGCGAGACTCTATCTCAAAAAAAT	1815
Db	3353	CATGGTTGGTCCCACTGCACCTTCAGCCCTGGGTGACAGAGTAAAGCCCTGTCTAAAAAGCA	3412
Qy	1816	AAATAAATAAAATAAAGGATCCGAGAGAGAAACAA	1848
Db	3413	AATAAGATGCAGTTTGGAGTGTAAATAGAGAA	3445
RESULT 3			
HSM803026			
LOCUS			
DEFINITION	Homo sapiens mRNA; cDNA DKFZp76100217 (from clone DKFZp76100217).		
ACCESSION	AL7113681		
VERSION	AL7113681.1 GI:19584382		
KEYWORDS	HTC.		
SOURCE	Homo sapiens (human)		
ORGANISM	Homo sapiens		
	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini; Hominidae; Homo		
REFERENCE	1 (bases 1 to 3990)		
AUTHORS	Ansoorge,W., Krieger,S., Regiert,T., Rittmueller,C., Schwager,B., Mewes,H.W., Weill,B., Amid,C., Osanger,A., Fobo,G., Han,M. and Wiemann,S.		
CONSRMT	The German cDNA Consortium		
TITLE	Direct Submission		
JOURNAL	Submitted (22-SEP-2004) MIPS, Ingolstaedter Landstr.1, D-85764		
COMMENT	Neuberberg, GERMANY Clone from S. Wiemann, Molecular Genome Analysis, German Cancer Research Center (DKFZ); Email s.wiemann@dkfz-heidelberg.de; sequenced by EMBL (European Molecular Biology Laboratories, Heidelberg/Germany) within the cDNA sequencing consortium of the German Genome Project. This clone (DKFZp76100217) is available at the RZPD Deutsches Ressourcenzentrum fuer Genomforschung GmbH in Berlin, Germany. Please contact RZPD for ordering: http://www.rzpd.de/cgi-bin/products/cl.cgi?CloneID=DKFZp76100217 Further information about the clone and the sequencing project is available at http://mips.gsf.de/projects/cdna/ .		
FEATURES	Location/Qualifiers		
	1..3990		
source	/organism="Homo sapiens"		
	/mol_type="mRNA"		
	/db_xref="RZPD:DKFZp76100217"		
	/db_xref="taxon:9606"		
	/clone="DKFZp76100217"		
	/tissue_type="amygdala"		
	/clone_lib="761 (synonym: hamy2). Vector pSport1; host DH10B;_sites NotI + SalI"		
	/dev_stage="adult"		
	/note="cytochrome c"		
	1..3990		
	/gene="DKFZp76100217"		
	61..378		
Gene	/gene="DKFZp76100217"		
	/codon_start=1		
CDS	/product="hypothetical protein"		
	/protein_id="CAD28485.1"		
	/db_xref="GI:19584383"		
	/db_xref="GOA:P99999"		
	/db_xref="InterPro:IPR002327"		
	/db_xref="InterPro:IPR003088"		
	/db_xref="UniProt/Swiss-Prot:P99999"		
ORIGIN	/translation="MGDVEKGKKIFIMKCSQCHTVTEGGKKHTQPNLHGLFGRITGQA		
	PGYSTAANKNGIHWGSDITLMEYLENPKYIPGTVMIFVGIKKCERADLIAYLKKA		
TNE"			

ORIGIN

```
Query Match      17.1%; Score 341; DB 4; Length 3990;
Best Local Similarity 57.5%; Pred. No. 1.1e-21;
Matches 856; Conservative 1; Mismatches 541; Indels 91; Gaps 10;

QY 404 TTTTCTTTTGTGTTTGTAGAGACAGAGTCTTGCTCTGTACACCGAGCATGAGCACAGT 463
DB 2522 TTTTCTCATTATTTTCTGAGCGGAGTTTGGCCCTTTTCGCCAGGCTGAGTACAGT 2581
QY 464 GGTGCAACATAGGTCACTGACGAGCTCAACCTCTGAGCTCAAGGGATCTGCTGACCTCA 523
DB 2582 GGTGCAATC-----TCACTGCAACCTCGCCCTTCAGTTGCAAGTGAATCTCTGCTCA 2636
QY 524 GCTCTCCCAAGTAGCTGGGACTACGAGCGTGCACACAGCGCTGCTGCTTAATAAAAAATT 583
DB 2637 GCTCTCTGAGTAGCTGGGATTGCAAGCACCCGCGCCACCGCTGCTGCTTAATTTTGTGA--T 2694
QY 584 TTTTGTGAGAGACTGGTCTTACTAGCTTGGCGAGCTTGTCTTAACTCTCTGGCTTCAA 643
DB 2695 TTTTGTGAGAGTGGTGTTCACCATGTTGGCAGACTGCTTGGACTCTCTGACTC-- 2752
QY 644 GCAATCCTCTACCTTGGCATCCCAAGTGTCTGGGATTAACAGGGGTGAGCCACCATGTC 703
DB 2753 GTGATCCACCCACCTTGGCTCCCAAGTTCTAGGATTACAGTGTGAGGCCACCAAGTCC 2812
QY 704 GGTCTACTTATT----- 714
DB 2813 GGCCCAATTTTCTCATTTTCTATGCTCTCTATTAAGGTCTGTGTTGGCACAGATGAGTA 2872
QY 715 -----TCCTTACATCCATCTTCCATAGATGTAGATCCACAGACAGGGATTACT 768
DB 2873 ACTGCCAATGTTCTAGTCAAGTATACCAAGCACTTCTGGTGGTTTAAATGTGATTCG 2932
QY 769 GCTATTTTCTCTCTTTCTTTTTCAGACAGAGTCTCACTTCATCAAC----- 816
DB 2933 TAACTTTTATTTTATTTTATTTTGTAGATAATTTCACTCTGTGTGCCAGGCTGGAGTGC 2992
QY 817 -TCAACCTCGGTTAGCTCACTGCAACCTCTGCTCCCGGGTTCAAGYATTTCTCTGCC 875
DB 2993 AATGGCGTGATGGCTCTCACGCAACCTCCGCTTCCAGGTTCAAGCGATTTCTCTGAC 3052
QY 876 TAAGCTCTCTAGTAGCTGGAATTAACAGCGTGCACCACTGCTTGGCTTAATTTTGT 935
DB 3053 TCAGCTCTCAAGTAGCTGGATTAACAGATGCGGCACATGCCAGCTTATTT--TGT 3110
QY 936 ATTTTTCAGCAGATGGGGTTTACCATGTTGCCAGGCTGGTCTCAAACTCTGACCTC 995
DB 3111 GTTTTTCAGACAGAGGTTTCTCCATGCTGCAGAGGCTGGTCTTGAATCTCCGAGCTC 3170
QY 996 AAGTAGTCTGCTGCTCACTGCTCCCAAGTGTGGAATTAAGGCTGAGTCACTGTGC 1055
DB 3171 GGGTGATCTGCTGCTCGCTCCCAAGTGTGGGATTAACAGGCTGAGCCACCAAGC 3230
QY 1056 CTGCGCGATTAAGTCTATTTTCTTATTTCTATATCCAGATCTAGAGCAGTGTCTCA 1115
DB 3231 CTGCGCAATTA---TGTAAATTTTAAAGGACATTTCTATCAGGGATATATACCTTCA 3287
QY 1116 CATATAGTAGGTGCTCAATAAATAATGATGAATGCACAGCTAGATATAAACTTTCTTT 1175
DB 3288 GAATAAGGAATAAGGGAAATAAGAGCACTATAAACCACATGTTTTCATTCTAGTGC 3347
QY 1176 TCTTTTAAACAATCTGTGACACTTTCAGATTAATAATAAATAAATAAATAAATAAATAA 1235
DB 3348 TCTGCTAAGTGGCTAGGTGTTAGTAAATCAAAACCAAGGGCCAGATGATTTTAAAGGGT 3407
QY 1236 TTTCACTATCACTGTTATGACTTTTTCATATGCTCAACCTTTATTTGTTACTGTT 1295
DB 3408 ATTCAAGATGCCACTACATGCTTATTTGTCTAGACAGTGTCTGTCTAATAGAACTTCT 3467
QY 1296 TTTTCAATGTTACTATTTTATGCTCACTGAATAAATAATGCTTTAAATTTGCTTATACATCTCC 1355
DB 3468 GTGACGATGGATATTTTGTAGACTTT-----TGCTGTCCAGTGTGGTACCCACTAACCA 3521
QY 1356 TGCTCCACTTTAGAGGCCAAATTTACAAATCTGATGAAGCTATGAACCTCTCCCCAG 1415
```

```
DB 3522 CATGTGCTGTAAAGCCCTTGAATAATATAGTGTGACTAGAAAGTATTTTATTTTAAA 3581
QY 1416 AGAATAACACACACACACTCAACAGTGTTCCTTTTAACTTTTTCACACTAAGA 1475
DB 3582 TTTACATAGGCACAAAGTGGCTAGTGGCTACTGTATTGACATTTCTGGGTCTTAGGACTAGAA 3641
QY 1476 CAAGAAACCTTCATTAGAG--ATGTTTGTTCATPATAATTTAAAAATAAATCACTCAGTGTGGC 1533
DB 3642 CCAGTGTGTGTAAACAAAGTACTTCTCTTTACTCTATTAAATCTAGAAATAGCCGGC 3701
QY 1534 ACAGTGAATCTCAAGCTCTTAACCAACAGTACTTTTGAAGTCCAAAGTGGGTGGATCACTTGA 1593
DB 3702 ATGCTGCTCATGCTGTGGTCCAGCACTTTGGAGGCCCAAGCGACGATCACTTGA 3761
QY 1594 GGTGAGAAAGTTCGAGACAGCTGGTCAATATGTTGAAACCTTCTCTACTATAAATAATAC 1653
DB 3762 GGTGGGCGTTTGAGACCAAGCTGGTCAACATGGCAACCTCTCTCTACAAAAACAT 3821
QY 1654 AAAAATTAGCTGGGTGTAGTGCATGCTCTGTAGTCCAGTACTCTCGGAGGCTGAGGC 1713
DB 3822 AAAAGTTAGCCAGGTGTGGTGGGCACTGTGTAGTCTCAGTACTTGGGAGGCTGAGGC 3881
QY 1714 AAGAGAATTGCTTCAACCTGGGAGGCGAGGTTTGCAGTGAGCCGAGATPCCCACTCA 1773
DB 3882 ACAAGATCACTTGAACCTTGGAGGTGAGGTTCAGTGAGCCAGGATTTGTGCCACTGCA 3941
QY 1774 CTCAGCCTGGGCGACACAGCGAGACTCTATCTCAAAAAATAATAATA 1822
DB 3942 CTCAGCCTGGGTGAC--GAGTGAACCTGTCTCCAAAAAATAAAAAA 3989

RESULT 4
AQ839852/c
LOCUS
DEFINITION
260L13-C56 CITB Homo sapiens genomic clone 260L13, genomic survey
sequence.
ACCESSION
AQ839852
VERSION
AQ839852.1 GI:6652484
KEYWORDS
GSS.
SOURCE
Homo sapiens
ORGANISM
Homo sapiens (human)
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini;
Homnidae; Homo.
REFERENCE
1 (bases 1 to 15970)
AUTHORS
Carpten,J.D., Makalowska,I., Robbins,C.M., Scott,N., Sood,R.,
Connors,T.D., Bonner,T.I., Smith,J.R., Faruque,M.U., Stephan,D.A.,
Pinkett,H., Morgenbesser,S.D., Su,K., Graham,C., Gregory,S.G.,
Williams,H., McDonald,L., Baxeianis,A.D., Klingler,K.W. and
Landes,G.M.
TITLE
A 6-Mb high-resolution physical and transcription map encompassing
the hereditary prostate cancer 1 (HPC1) region
JOURNAL
Genomics 64 (1), 1-14 (2000)
PUBMED
10708513
COMMENT
Contact: Carpten JD
Cancer Genetics Branch
National Human Genome Research Institute/National Institutes of
Health
Bldg. 36, Room 3D04, 36 Convent Drive, Bethesda, MD
Tel: 301 435 5626
Fax: 301 435 5465
Email: jdc@nhgri.nih.gov
Class: shotgun.
Location/Qualifiers
1..15970
/organism="Homo sapiens"
/mol_type="genomic DNA"
/db_xref="taxon:9606"
/clone="260L13"
/clone_lib="CITB"

FEATURES
source
```

ORIGIN

Query Match	16.9%;	Score 337.2;	DB 9;	Length 15970;
Best Local Similarity	57.8%;	1; Mismatches 529;	Indels 131;	Gaps 12;
Matches 905;	Conservative			
QY	398	TTTGGGTTTTTTTTTTTGGTTTGTGTTTAGAGACAGGGCTCTTGCTCTGTCACCCAGGCGATGAG	457	
DB	6787	TTTTTATTTTTTATTTTTTTTTTTTTTTGAGGTGGAGTATCGCTATGTCACCCAGGCTCGAG	6728	
QY	458	CACAGTGGTGCAACCATAGGTCACCTGCAGCCCTCAACCTCTCTGAGCTCAAGGGATCTGCTG	517	
DB	6727	TGCAGTGGCATGATACAGCTCACTGCACAACTTCACCTGCTAGATTCAAGCAGTCTCTCGT	6668	
QY	518	ACCTCAGCCTCCCAAGTAGCTGGGACACTAGAGCGTGCAACCAACACGAGCTGGCTAATTTAA	577	
DB	6667	GACTCAGCCTCCTGAGTAGCTGTGATTAACGGCAACATGCCACCAATGCTACGCTAACT-TT	6609	
QY	578	AAAAATTTTTTTGTAGAGACTGGGTCCTTACTACGTTTGGCCAGGCTGTGCTTTAAACTCTCTGG	637	
DB	6608	TGTATTTTTTAGTAGACAGGGTCTCACATGTTTGGCCAGGCTGGTCTGGAACTCTCTGA	6549	
QY	638	CTTCAAGCAATCTCTCTACCTTGTCATCCCAAGTGTCTGGGATTTACAGGGGTGAGCCACC	697	
DB	6548	CCTCAAGTGATCCGCCCTCTCTGGCCTCCCAAGTGCTGGGATTTACAGGCATGAGCCACT	6489	
QY	698	ATGTGGCGGTACTTATTTCTTTT-----ACATTCCATCTTTTCCAAATAGATGTAAAGAT	749	
DB	6488	CGCGCCAGCCTCTGAGTCTTTTTAAGTCAGAAACTTGAGCTCTTCAGCTCTGTATAAATTT	6429	
QY	750	CCACAGAACAGGGATTAATCGCT--ATTCTTCTCTTCTTTTGTAGACAGAGTCTCAC	807	
DB	6428	GGGGGGCAGGGGACTAAATTTTTTCTTTTCTTTTCTTTTAAAGATGGAGTCTGTG	6369	
QY	808	TTCATCACCTCAACCTCCGTTCA-----GCTCACCTGCACAACTCTGCTCTCCC	853	
DB	6368	TGCTGTGCCAGGGTGGAGTGCAGTGGTGATCTTGGCTCACTGCAACTCTGCTCTCC	6309	
QY	854	GGGTTCAAGTGATTTCTCTGCTTAAGCCTCTGAGTAGCTGGAAATTAACAAGCTGCACCA	913	
DB	6308	AGGTTCAAGCAATTTCTCTGCTCAGCTTACTGAGTAGCTGGGACTCGAGGCTGTGCCA	6249	
QY	914	CCATGCTTGGCTAAATTTTTTGTATTTTTTACAGAGATGGGGTTTTACATGTTGCCAGG	973	
DB	6248	CCACTCTCAGCTAA-TTTTTGTATTTTTTAGTAGACAGGGTTTCAGCAATTTGGCCAGG	6190	
QY	974	CTGCTCTCAAACTCCTGACCTCAAGTGATCTGCTGCTCAGTCTCCCAAGTGCTGGAA	1033	
DB	6189	CTGCTCTCGATCTCCTGACCTCAAGTGATCTGCTACCTGCTCCCAAGTGCTGGGA	6130	
QY	1034	TTATAGCGGTGAGTCACTGTGCTGGCGGATTAATGTCATAATTTCTTTATGCTATATCC	1093	
DB	6129	TTACAGCAACAGGCCACACACCTGGCCCTT--GGGGGATGTTATTTCTTTTGACAAGTT	6073	
QY	1094	CCAGATCTAGAGCAGTCTGACATATAGTAGGTGCTCAATAATAATTGATCAATGTCAC	1153	
DB	6072	TTGCCITTCAAATTAATATCTGTGTCCTCTTCAGGAACTCTGTTTATTAATTTTGGGT	6013	
QY	1154	AGCTAGATATAAACTTTCTTTTCTTTTAAAACAACTTGCACAACTTTTCGCAATA	1213	
DB	6012	CTTCTAGATAATCCCTTAATTTTTTAAATAATCTATACGGTTCACTCTTTGGCAATTA	5953	
QY	1214	AATACAATCTTGCAATCTGCTTTTCACTTATACCTTGTTATGACTTTTTCATATGCC	1273	
DB	5952	GTTCTACTTTATACTTTTTCCTTAAATTTTATTTTCCAACTCTTAATTTAAATTTCTGTCA	5893	
QY	1274	TCAAACTTTATTT-----GTTACTGTTTTTTTCAATGTTACTATTTTAGTC----	1318	
DB	5892	TATTTTGTTCATTCTTAAGAGTTATTTTCATATTTTTTCACTGTCTCTTTTCTTTT	5833	
QY	1319	-----ACTGAATAATATGGCTTAATTTGCTTATATACATCTCTGCTCCACTTTTGA	1369	
DB	5832	GGCTAGTCAAGTGAATATGTTCTTTTTTAATAGTGTCAATATGTTTCAGGATACAA	5773	
QY	1370	A-----GGCCAAATTTTACAATCTGATGAAGC	1397	


```
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/clone="IMAGE:30390116"
/lab_host="DH10B-Ton A ( T1 and T5 phage resistances)"
/clone_lib="NIH_MGC_180"
/note="Organ: Testis; Vector: pCMV-SPORT6.1; Site.1: NotI; Site 2: EcoRV (destroyed); Library is oligo-dT primed and directionally cloned (EcoRV site is destroyed upon cloning). Average insert size 1.68 kb. Library was constructed by (Invitrogen). Note: this is a NIH_MGC library."
```

```
ORIGIN
Query Match      15.8%; Score 315.8; DB 6; Length 897;
Best Local Similarity 68.1%; Pred. No. 4.4e-19;
Matches 469; Conservative 1; Mismatches 213; Indels 6; Gaps 2;

QY 383 TGTAAAGTGGTGACAGTTTGGGTTTTTTTGTGTGTTTGTAGACAGAGGTTCTGCTCTG 442
Db      |||||
QY 765 TGTATGCATCCGATATTATTATTTATTTATTTTATTTTGGAGACATGGTCTCACTCTG 706
Db      |||||
QY 443 TCACCCAGCATGACACAGTGTGCACCATAGTCTACAGCTCAAGCTCAAGCTCTGAGC 502
Db      |||||
QY 705 TTGCCCCAGCTGGAGTGCAGTGGCAGGATCTCAGCTCACTGCAACTCCACCTCCAGGT 646
Db      |||||
QY 503 TCAAGGGATCTGCTCACTCTCAAGCTCCCAAGTAGCTGGGACTACGAGGCTGCACCAACCAC 562
Db      |||||
QY 645 TCAGCAATCTCCACCTCAAGCTCCCAAGTAGCTGGGATACAGGTGTGCACCAACCAC 586
Db      |||||
QY 563 GCCTGGCTAAATTAATAAATTTTTTGTAGAGACTGGGTCTTACATGTTGGCCAGGCTT 622
Db      |||||
QY 585 GCCCAGCTAAATTTTTTGTATTTAAGTACAGATGGGTTTTTGCCTAGTGTGGCCAGGCTC 526
Db      |||||
QY 623 GTCTTAACTCTGGCTTCAAGCAATCTCCTACCTTGGCATCCCAAGTGTCTGGGATTA 682
Db      |||||
QY 525 CTCTCAATTTCTGACCTCAAGTATCTGCCCACTTGGCCCTTCCCAAGTGTGGGATTA 466
Db      |||||
QY 683 CAGGGGTGAGCCACCATGTGGCGTACTTATTTTATCATTTCCATCTTTTCAATAGAAAT 742
Db      |||||
QY 465 CAGGCGTGAGCCACCGTGGCCAGCTGCATCTGTATTTAATGCTTGCCAGAAATTTGATT 406
Db      |||||
QY 743 GTAAGATCCACAGAACAGGGATTACGCTATTTCTTCTCTTTCTTTTGTAGACAGAT 802
Db      |||||
QY 405 GCACCATGCTTATTGGCTTCTATCTTTTCTTTTCTTTTGTAGATGGAGACTCACTCTCT 346
Db      |||||
QY 803 CTCACCTTCACCTCAACTCGT---TCAGCTCACTGCAACCTCTGCTCCCGGTT 858
Db      |||||
QY 345 CTCCAGGCTGGAGTGCAGTGGTGTGATCTCAGCTCACTGCAACCTTCCACTTCTCTGGGT 286
Db      |||||
QY 859 CAAGYGATTTCTCTGCTTAAGCTCTGAGTAGCTGGAATTAACAAGCGTGCCACCATG 918
Db      |||||
QY 285 CAAGCGATTTCTCTGCTTACGCTCCAGAGTAGCTGGGACTACAGATGTGTGCCACCA 226
Db      |||||
QY 919 CTTGGCTAAATTTTGTATTTTGTAGCAGAGATGGGTTTTTACATGTGCGCCAGGCTGGT 978
Db      |||||
QY 225 CTTGGCTAA--TTTCGTAATTTTGTAGTAGAGATGGGTTTCCACCATGTGTGCCAGACTAAT 168
Db      |||||
QY 979 CTCAAATCTCTGACCTCAAGTATCTGCTGCTCAGTCTCCCAAGATGCTGAATTA 1038
Db      |||||
QY 167 CTCGAATCTCTGACCTCAGGTATCTGCTGCTGCTGCTTCCCAAGATGCTGGGTTTACA 108
Db      |||||
QY 1039 GCGGTGAGTCACTGTGCTGGCCGATTAAC 1067
Db      |||||
QY 107 GCGGTGAGCCACCAACCCAGCGGCTTCC 79
Db      |||||
```

```
RESULT 6
CR858326
LOCUS      3095 bp      mRNA      linear      HTC 12-NOV-2004
DEFINITION Pongo pygmaeus mRNA; cdna DKFZp469P012 (from clone DKFZp469P012).
ACCESSION CR858326
VERSION   CR858326.1 GI:55727613
```

```
KEYWORDS      HTC.
SOURCE        Pongo pygmaeus (Orangutan)
ORGANISM      Pongo pygmaeus
              Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
              Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini;
              Homiidae; Pongo.
REFERENCE     1 (bases 1 to 3095)
AUTHORS      Bloeker,H., Boecher,M., Brandt,P., Mewes,H.W., Weil,B., Amid,C.,
              Oanger,A., Fobo,G., Han,M. and Wiemann,S.
              The German cDNA Consortium
CONSTRM      Direct Submission
TITLE         Submitted (12-NOV-2004) MIPS, Ingolstaedter Landstr.1, D-85764
              Neuherberg, GERMANY
COMMENT       Clone from S. Wiemann, Molecular Genome Analysis, German Cancer
              Research Center (DKFZ); Email s.wiemann@dkfz-heidelberg.de;
              sequenced by GBF (National Research Centre for Biotechnology Ltd.,
              Braunschweig/Germany) within the cDNA sequencing consortium of the
              German Genome Project.
              This clone (DKFZp469P012) is available at the RZPD Deutsches
              Ressourcenzentrum fuer Genomforschung GmbH in Berlin, Germany.
              Please contact RZPD for ordering:
              http://www.rzpd.de/cgi-bin/products/cl.cgi?cloneID=DKFZp469P012
              Further information about the clone and the sequencing project is
              available at http://mips.gsf.de/projects/cdna/.
FEATURES      Location/Qualifiers
              1..3095
               /organism="Pongo pygmaeus"
               /mol_type="mRNA"
               /db_xref="taxon:9600"
               /clone="DKFZp469P012"
               /tissue_type="kidney"
               /clone_lib="469 (synonym: pkidi). Vector pSport1_Sfi; host
               DH10B; sites SfiIA + SfiIB"
               /dev_stage="adult"
               /note="hypothetical protein (Homo sapiens)"
              1..3095
               /gene="DKFZp469P012"
              120..533
               /gene="DKFZp469P012"
               /codon_start=1
               /product="hypothetical protein"
               /protein_id="CAH90562.1"
               /db_xref="GI:55727614"
               /translation="MISGRVPLVSMAGRVCLCGSAGSAGVAVVEAIIKRLQAL
               SPEVLELRNKGSHVPPGSETHFRVAVVSRFELSLPQRHLRIHAALABELAGPVH
               ALATQARTPAQWGENSLDTPSPCLGGNKKTLGTP"
ORIGIN
Query Match      15.6%; Score 311; DB 4; Length 3095;
Best Local Similarity 56.6%; Pred. No. 5.2e-19;
Matches 883; Conservative 1; Mismatches 511; Indels 166; Gaps 10;

QY 384 GTAAGTGGTGACAGTTTGGGTTTTTTTGTGTGTTTGTAGACAGAGGTTCTGCTCTG 443
Db      |||||
QY 1570 GTGAGCCACTACTCTAGCGTTTTTGTGTTTGTGTTTGTAGAGTGGAGTCTCACTTTGT 1629
Db      |||||
QY 444 CACCCAGGCAVAGACAGTGTGTGCACCAATAGGTCACTGACGCTCAACCTCTGAGCT 503
Db      |||||
QY 1630 TGCCCAACCTGGAGTGCAGTGTGGGATCTCAGCTCACTGACGCTCCCGCTCCCAAGTT 1689
Db      |||||
QY 504 CAAGGATCTGTGACCTCAGCTCCCAAGTAGCTGGGACTACGAGCGTGCCACCAACG 563
Db      |||||
QY 1690 CAAGCAATTTCTCTGCTCAGCTCTCCCAAGTAGTGGGACTAGAGCTTTGTGCCACACA 1749
Db      |||||
QY 564 CTGCGCTAATTAATAAATTTTTTGTAGAGACTGGGCTTACTACTGTTGGCCAGGCTTG 623
Db      |||||
QY 1750 CCCAGCTAAT--TTTGTATTTTATGGAGACAGGTTTCCACATGTTGCCAGGCTGG 1806
Db      |||||
QY 624 TCTTAACTCTCGGCTTCAAGCAATCTCTCACTTGGCATCCCAAGTGTCTGGGATTAC 683
Db      |||||
QY 1807 TCTTAGACTCTGAGCTCAGGCACCTCCGCCACCTCCAGGCTCCCAAGGTGCTAGGATTAC 1866
Db      |||||
QY 684 AGGGGTGAGCCACCACATGTGCGGCTACTTATTTCTTTACATTC----- 725
```

[illegible]

Db 2865 CTGTAATCCCGAGCTCTTAGCGAGCTGAGCGACGAGAAATCACTTGAACCTGGGAGGCAGA 2924

QY 1743 GGTTCAGTGCAGCGAGATCCACCACTGCACCTCCAGCTCCGCGCGACACAGCGAGACTCT 1802

Db 2925 GGTTCAGTGCAGCGAGATCAGCGCACTGCACCTCCAGCTCCGCGCGACCAAGTGAACCTCT 2984

QY 1803 ATTCACAAAAATAAATAAATAAAGATCGGAGAGAAAAACAAACATAAATAAGATTC 1862

Db 2985 GTCTCAAAAAAATAAATAAATAAAGATCGGAGAGAAAAACAAACATAAATAAGATTC 3044

QY 1863 C 1863

Db 3045 C 3045

RESULT 7

CR749233/C

LOCUS CR749233 Homo sapiens mRNA; cDNA DKFP686C0331 (from clone DKFP686C0331). 5797 bp mRNA linear HTC 19-AUG-2004

DEFINITION CR749233 Homo sapiens mRNA; cDNA DKFP686C0331 (from clone DKFP686C0331).

ACCESSION CR749233

VERSION CR749233.1 GI:51476197

KEYWORDS HTC.

SOURCE Homo sapiens (human)

ORGANISM Homo sapiens

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini; Hominiidae; Homo.

REFERENCE 1 (bases 1 to 5797)

POUSTKA A., ALBERT R., MOOSMAYER P., SCHUPP I., WELLENREUTHER R., MEWES H.W., WEILL B., AMID C., OSANGER A., FOBO G., HAN M. and WIEMANN S.

AUTHORS The German cDNA Consortium

CONSTRM Direct Submission

TITLE Submitted (17-AUG-2004) MIPS, Ingolstaedter Landstr.1, D-85764

JOURNAL Neuherberg, GERMANY

COMMENT Clone from S. Wiemann, Molecular Genome Analysis, German Cancer Research Center (DKFZ); Email s.wiemann@dkfz-heidelberg.de; sequenced by DKFZ (German Cancer Research Center, Heidelberg/Germany) within the cDNA sequencing consortium of the German Genome Project. This clone (DKFP686C0331) is available at the RZPD Deutsches Ressourcenzentrum fuer Genomforschung GmbH in Berlin, Germany. Please contact RZPD for ordering: <http://www.rzpd.de/cgi-bin/products/cl.cgi?Cloneld=DKFP686C0331> Further information about the clone and the sequencing project is available at <http://mips.gsf.de/projects/cdna/>.

FEATURES

source Location/Qualifiers

1..5797

/organism="Homo sapiens"

/mol_type="mRNA"

/db_xref="taxon:9606"

/clone="DKFP686C0331"

/tissue_type="testis"

/clone_lib="686 (synonym: hlcc3). Vector pSport1_sfi; host DH10B; sites SfiIA + SfiIB"

/dev_stage="adult"

/note="hypothetical protein"

1..5797

/gene="DKFP686C0331"

103..1725

/gene="DKFP686C0331"

/codon_start=1

/product="hypothetical protein"

/protein_id="CAH18089.1"

/db_xref="GI:51476198"

/translation="MGPIQFRDVAIEFLEEWHCLDTAQRNLRYNMULENTSNLVFLG
ITVSPDLITCLEQGRKPLTKRMENIAKPSVMCSHFQADLWPEQSMKDSFKQVVLRR
YFKCBHDNLQLKQKSVDECKVKEGYNELNQCLTTPRIQCQDKTVKVLHQPPNS
NGQKRGHTKPKFYIECGAKFKQPSLTLLTHKKIHTGGKPKYKCECGKAFNHSCSLTR
HKKIHTGKPKYKCECGKAFKHSSTLTTHKRNHTGKPKYKCDKCGKAFMSSTLSKHE
IHTGKPKYKCECGKAFNRSSSTLTTHKLIHTGKPKYKCECDKAFKFSYTLTHKRI
HTEDPKYKCECGKAFKFSSTLTTHKRIHTGKPKYKCECGKAFKRSDDLTHKRIHT
GSKPKYKCECGKAFKFSSTLTTHKRIHTGKPKYKCECGKAFKFSSTLTTHKRIHTG
KPKYKCECGKAFKFSSTLTTHKRIHTGKPKYKCECDKAFKHSSTLTTHKRIHTGKPK

ORIGIN		YRCBEGKAPKHSKLTTHKRFHTGKPYRCE"	
Query Match		15 4%; Score 307.2; DB 4; Length 5797;	
Best Local Similarity		56.9%; Pred. No. 7.4e-19;	
Matches 902; Conservative		1; Mismatches 569; Indels 114; Gaps 14;	
Qy	405	TTTTTTTGTGTTGTTTGTAGACAGAGGTCCTGCTGTCTCACCAGGATGACAGCAGTG	464
Db	4976	TTTTTTCTTTTTTTTTTGTGACAGAGTCCTGCTGTGCTGCTGCCAGGCTGGAGTGCAGTG	4917
Qy	465	GTGCAACATAGTCACGTGACGACCTCAACCTCCTGAGTCAAGGATCTGTCGACCTCAG	524
Db	4916	GCAGGATCTGGCTCACTGCAACCTCTGCTCTCTGGGTGTACACATCTCCCG-CCTCAG	4858
Qy	525	CCTCCCAAGTAGCTGGGACTACGAGCTGCACACACGCTGCTAATTAATAAAAAATTT	584
Db	4857	CCTCTGGAGTAGCTGGGACTACAGGCGCCACACCGCCAGCTAATTTA-----TT	4804
Qy	585	TTTTGTAGAGACTGGGTCTTACTAGCTTGGCGAGGCTTGTCTTAAACTCTCGGCTCAAG	644
Db	4803	TTTAGTAGACAGAGGTTTCAACGTGTTAGCCAGGATGCTTGTGATCTCTCTGACCTCATG	4744
Qy	645	CAATCCTCCTACTTGGGATCCAAAGTCTGGGATTTACAGGGTGACGCCACCATGTGCG	704
Db	4743	--ATCCACTGCTGGGCTCCCAAGTCTGGGATTTACAGGTGTGAGCCACTGTCCCTG	4686
Qy	705	GCTACTTATTTCTTTACATTCATCTTCTCCATAGATGTAGATCCACAGAACAGGAT	764
Db	4685	GCTTTTTTCTTCTT-----TTTGAGATGGAGTGTCCCTCTGT	4648
Qy	765	TACTGCTATTTTCTTCTTCTTTTGTGAGACAGAGTCTCATCTTCATCACTCAACCTC	824
Db	4647	CACCCAGGCTGAGTGCAGTGTGCTAT-----	4620
Qy	825	CGTTCAGTCTCATGCAACTCTGCTCCCGGTTCAAGTGATCTCTCCGCTCAAGCTCC	884
Db	4619	--CTCAGCACACTGAACCTCCACCTCCAGGTTCAAGCAATCTCTGCTCCGCTTCC	4562
Qy	885	TGAGTAGCTGGAATACAGCTGCACACCATGCTTGCTTAATTTTGTATTTTATAGC	944
Db	4561	CGAGTAGCTAGGAATAC--GGGTGCACACCATGCCCCAACTAATTTTGTGTTT-----	4508
Qy	945	AGAGATGGGTTTACCATGTTGCCAGGCTGGTCTCAAACTCTGACCTCAAGTGAATC	1004
Db	4507	-AGTAGAGATTTTCAATGTTGGTCAGGCTGGTCTTGAACCTCTGACCTCA--TGATCT	4451
Qy	1005	GCTGCTCAGTCTCCAAAGTCTGGAATTTAGCGGTGAGTCACTGTGCTGGCCGAT	1064
Db	4450	GCCGACGTCACCTACCAAGTCTGGGATTTACAGGCATGAGCCACTGGCTCCAGCCCA-	4392
Qy	1065	TACTGTCTATTTCTTTATGCTATATCCAGATCTAGAGCAGTGTCTGACATAGTA	1124
Db	4391	-----TATTCTATTTATTTCTGTATAATTTTATATGACATATAAAATCACCTGTA	4337
Qy	1125	GGTGTCTCAATAATAATTTGATGATGACACAGCCTAGATATAAACTTTCTTTTCTTTT	1184
Db	4336	GTCAATTTACATTTAATTTGACATTTAAAG--TAACTAAAGTGTATAATTACACTGT	4279
Qy	1185	TAAACAACTTGACAACTTTTGACAGATAAATCAATCTGCACTTCTGCTTTTTCATTA	1244
Db	4278	TGGAATAGAAGGATAAATCTAGAGGTGATAGATACCTTATTTACCTTAATGTAATTAC	4219
Qy	1245	TCACCTTGTATGACTTTTTCATATTG-----CCTCAAACTTTTATGTTTACTGTTTTT	1299
Db	4218	TACATATTTGCTGGAATCAAAATTTCTTATAGACATATATACATFACTATATA	4159
Qy	1300	CATTGTTACTATTTTACTGACTGAATAATATGGCTT---AATTGGCTTATACATCTCTCT	1356
Db	4158	CTCACAATAGTAAATAATTAAGAAAGAAAGATTAATAAATTTAACTAACGGGAACAAT	4099
Qy	1357	GCTCCACTTTAGAGGCCAAATTTACAAATCTGTATGAAGCTATGAACCTCTCTCCACGA	1416

Db	4098	ATTCTTTCTTTGCGAGTTTACTGGCAAGTGATTACTAGTGATGACATTCACATACATACC	4039
Qy	1417	GAATACACACACACACACTCACACACAGTTTTTTTTTAAATG-----	1462
Db	4038	AAATAGTATATCTTTCTTCCATCTTTTGTCTTACACCATTTGATTAAGCCGATAGGTAAA	3979
Qy	1463	-----TTTGCACACTAAGACACAGAAACCTGCAATTAGAGGATGTTTGTTCATATTAATAA	1516
Db	3978	GTTAGTGGCGTAATATGCTTCATTAATGACACATAGTTTTTAAGATGGCAAAAATAAA	3919
Qy	1517	AAATAAATCAGTTAGTGGGACAGTCAACAGCCTGTAAACACAGTACTTTTGGAGTCCAG	1576
Db	3918	ATAAATTTAGGCTGGACATATTTGGCTCACACATATATCCACACACTTTGAGAGCCGAG	3859
Qy	1577	GTGGTGGATCATTGAGGTGAGAGTTTCAGACCCAGCCTGTCATATGTTGTAACCCCT	1636
Db	3858	GTGGGTGGATCGCTTGAGCTCAGGAGTTTGAGACTGACCTTGAACATATGGGAAACCCC	3799
Qy	1637	ATCTCTACTAAAAATACAAAAATTTAGCTGGGTGTAGTGATGCGATGCTGTAGTCCAGCT	1696
Db	3798	ATGTCTATTAATAATATAAATTTAGCCAGCGTGTGTCAAGCACCTGTAGTCCCACT	3739
Qy	1697	ACTCGGAGGCTGAGGCAAGAGAAATTCCTTTGAACTCTGGAGGCGAGAGTTGCAG-----TG	1752
Db	3738	ACTCGGAGGCTGAGGCGAGGAGAAATTTCTTGAACCTTGGAGGTGGAGGTTGCAGTCAATG	3679
Qy	1753	AGCCGAGATCCACCACTGCACCTCCAGCCTGGGCGACACAGCAGACTCTATCTCAAAA	1812
Db	3678	TGCTGAGACTGCATGCTGCTCCAGCCTGGGTGACAGAGCAAGACTTCCATCCCAAG	3619
Qy	1813	AATAAATAATAATAATAAAGGATCGGAGAGAAACAAACAACTAATAAGATTCCTGAAGTAA	1872
Db	3618	AAAGTTAAGTAATAATAATAATTAAGTTTCAACATATCTAACAATTTTAAATAT	3559
Qy	1873	GCAGAGATACGTAATAATATATGTAATAAGTTTAAATGCAATTTAACTGTATTTATG	1932
Db	3558	ATTGCATTTTATACGTAACAGTAAATTTAGTAAATAATTTGATATACTTTAATTTATAA	3499
Qy	1933	TTTATTTTGGTTATAAAGTAAACAA	1958
Db	3498	TTTCTTCTCTCAGTATAATGTAGAA	3473

RESULT 8
CR860521/c 1605 bp mRNA linear HTC 12-NOV-2004
LOCUS Pongo pygmaeus mRNA; cDNA DKFZp468P2419 (from clone DKFZp468P2419).
DEFINITION CR860521
ACCESSION CR860521.1 GI:55731876
VERSION CR860521.1
KEYWORDS HTC.
SOURCE Pongo pygmaeus (orangutan)
ORGANISM Pongo pygmaeus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini;
Hominoidea; Pongo.
REFERENCE 1 (bases 1 to 1605)
AUTHORS Ottenwaelder, B.; Obermaier, B., Deutschenbaur, S., Schaipe, A.,
Wewes, H.W., Weill, B., Amid, C., Osanger, A., Fobo, G., Han, M. and
Wiemann, S.
CONSRMT The German cDNA Consortium
TITLE Direct Submission
JOURNAL Submitted (12-NOV-2004) MIPS, Ingolstaedter Landstr.1, D-85764
Neuherberg, GERMANY
COMMENT Clone from S. Wiemann, Molecular Genome Analysis, German Cancer
Research Center (DKFZ); Email s.wiemann@dkfz-heidelberg.de; cDNA
sequenced by Medigenomix (Martinsried/Germany) within the cDNA
sequencing consortium of the German Genome Project.
This clone (DKFZp468P2419) is available at the RZPD Deutsches
Ressourcenzentrum fuer Genomforschung GmbH in Berlin, Germany.
Please contact RZPD for ordering:
http://www.rzpd.de/cgi-bin/products/cl.cgi?CloneID=DKFZp468P2419
Further information about the clone and the sequencing project is
available at http://mips.gsf.de/projects/cdna/.

```
FEATURES
  source
    Location/Qualifiers
      1..1605
        /organism="Pongo pygmaeus"
        /mol_type="mRNA"
        /db_xref="taxon:9600"
        /clone="DKP2p468P2419"
        /tissue_type="heart"
        /clone_lib="468 (synonym: phr1). Vector pSport1_sfi; host DH10B; sites SfiI + SfiIb"
        /dev_stage="adult"
        /note="unclassified"

ORIGIN
  Query Match      15.1%; Score 301.2; DB 4; Length 1605;
  Best Local Similarity 67.7%; Pred. No. 5.6e-18;
  Matches 474; Conservative 1; Mismatches 204; Indels 21; Gaps 3;

QY 398 TTGGGTTTTTTTTTTGTTGTTTGTAGAGACAGGGTCTGCTGTCTACCCAGGAGATGAG 457
Db 1602 TTTTTTTTTTTTGTGTTGTTGTTGTTGAGACAGAAATCTTGCTTCTCCCGGGCTGGAG 1543
QY 458 CACAGTGTGCAACCATAGGTCACTGACAGCTCAACCTCTGAGCTCAAGGAGATCTGCTG 517
Db 1542 TGTAGTGTGCAATCTAGCTCACTGCAACATCCATCTCCCGGATTCACATTTATCTCT 1483
QY 518 ACCTCAGCTCCCAAGTAGCTGGGACTAGAGCGTGCACACACCGCTGCTGCTAAATAAA 577
Db 1482 GCCTCAGCTCTCAGGTAGCTGGGATATGCGTCTGCCATAATGCCAACTCAATTTT 1423
QY 578 AAAAAATTTTTTTGAGAGACTGGTCTTACTAGTTGGCCAGCTTGCTTAACTCTCTGG 637
Db 1422 GTA--CTTTAGTAGAGACGGGGTTTTTACCAATGTTGGCCAGCTGGTCTCAAACTCTCTG 1365
QY 638 CTTCAAGCAATCTCTACCTTGGCATCCCAAGTGTGGGATTTACAGGGGTAGACCAACC 697
Db 1364 CTTCAAGTATCCACCTGCTGGCTCCCAAGTGTGGGATTTACAGGGGTAGACCAACC 1305
QY 698 ATGTGCGGCTACTTATTTTACATTCATCTTTTCCAAATAGAATGTAAGATCCACAGAA 757
Db 1304 CACCTGGCTGAAATTTCTCCATCTTCCCA-----CACACCGGCTCAGGTTCTCTCTCC 1250
QY 758 CAGGGATTAAGCTATTTCTCTTTCTTTTGTAGAGACAGATCTCACTTCACTCACT 817
Db 1249 TGATCATTTGACGCTTTTTTTTTTTTTTTTTTTTTTTTGTAGACAGCATCTCACTGTCTACCC 1190
QY 818 CAACCTCGTTCA-----GCTCAGTCAACCTCTGCTCCCGGGTTCAAGY 863
Db 1189 AGCTGGAGTGCAGTAGTGCAATCTTGCTCACTGAAACCTCTCTCCAGCCTCAAGC 1130
QY 864 GATTCTCTGCTAAGCTCTCTGAGTAGTGGAATTAAGAGCTGCACCACTGCTTGG 923
Db 1129 GATTCTCTGCTCAGCTCCGAGTAGCTGGGATTAATGTTGACCACTAACACCA 1070
QY 924 CTAATTTTTTGTATTTTGTAGAGATGGGGTTTTTACATGTTGCCAGGCTGGTCTCAA 983
Db 1069 GATAATTTTTTGTACTTTTGTAGAGATGGGGTTTTGACCATATTGGCCAGACTGGTCTTGA 1010
QY 984 ACTCTCAGCTCAAGTCACTGCTGCTGCTCAGTCTCCCAAGTCTGGAATTTAGGCGT 1043
Db 1009 ATTCTCAGCTCAGGTGATCGGCCACTCAGCTCCCAAGTGTGGGGTTATAGGCGT 950
QY 1044 GAGTCACTGTGCTGCGCGATTAAGTCTATTTTCTTTAT 1083
Db 949 GAGCCACTGCACCGCGCCCTTCTTCACTTAGTCAAT 910

RESULT 9
CR599842 1863 bp mRNA linear HTC 21-JUL-2004
LOCUS full-length cDNA clone CSODA002Y006 of Neuroblastoma of Homo
DEFINITION sapiens (human).
ACCESSION CR599842
VERSION CR599842.1 GI:50480649

KEYWORDS
SOURCE Homo sapiens (human)
ORGANISM Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 1863)
AUTHORS Li, W.B., Gruber, C., Jesses, J. and Polayes, D.
TITLE Full-length cDNA libraries and normalization
JOURNAL Unpublished
REMARK Contact: Feng Liang Email: fliang@lifetech.com URL: http://fulllength.invitrogen.com/ Invitrogen Corporation 1600 Faraday Avenue
REFERENCE 2 (bases 1 to 1863)
AUTHORS Genoscope.
TITLE Direct Submission
JOURNAL Submitted (20-JUL-2004) Genoscope - Centre National de Sequencage : BP 191 91006 EVRY cedex - FRANCE (E-mail: seque@genoscope.cns.fr)
COMMENT - Web : www.genoscope.cns.fr)
1st strand cDNA was primed with a NotI-oligo(dT) primer. Five prime end enriched, double-strand cDNA was digested with Not I and cloned into the Not I and EcoR V sites of the pCMVSPORT 6 vector. Library was normalized. Library was constructed by Life Technologies, a division of Invitrogen.
FEATURES
  source
    Location/Qualifiers
      1..1863
        /organism="Homo sapiens"
        /mol_type="mRNA"
        /db_xref="taxon:9606"
        /clone="CSODA002Y006"
        /tissue_type="Neuroblastoma"
        /plasmid="pCMVSPORT_6"

ORIGIN
  Query Match      15.1%; Score 301.2; DB 4; Length 1863;
  Best Local Similarity 56.9%; Pred. No. 5.1e-18;
  Matches 811; Conservative 1; Mismatches 509; Indels 105; Gaps 10;

QY 424 GAGACAGGGTCTGTGCTGTCCACCGAGATGACAGAGTGTGCAACCATAGGTCACTG 483
Db 496 GGGTGAAGTCTCACTCTGTTCCAGGCTGGAGTTCTGTGGCACCGTGTGGTCACTG 555
QY 484 CAGCTCTCAACCTCTCAGCTCAAGGGATCTGTGACCTCAGCTCTCCCAAGTGTGGGAC 543
Db 556 CAACCTCCGCTCTCTGGTTCAAGCAATCTCTTGTCTCAGCTCTCAGAGTAGTGGAT 615
QY 544 TAGAGGGTGACACACACCGCTGGCTTAATAAAAAATTTTTTTGTAGAGACTGGGTCT 603
Db 616 TAGAGGTGGCCACACACACCGGCTAATTTTTTTTATATTTTGTAGACACGGGGTTT 675
QY 604 TACTAGTGTGGCCAGGCTGTCTTAACTCTGGCTTCAAGCAATCTCTTACCTTGGCA 663
Db 676 CACCATGTTGGCCAGGCGGCTTTGAACCTGTGACGTGAGGTGATGACCGCTTGGCC 735
QY 664 TCCCAAAGTGTGGGATTACAGGGGTGAGCCACATGTGCGGTACTTATTTCTTTACAT 723
Db 736 TCCCAAAGTGTGGGATTACAGGTCCAGATTTTGAACAAATCTCGTAGAGATTGAACAG 795
QY 724 TCCATCTTTCCAAATAGAATGAAGATCCACAGAACAGGGATTAATCTGCTATTTTCTCT 783
Db 796 AGTGTGCGCTGGATCCCGCCACAGTAGTGAGGAGACCTGAAGACACAGAGAAAACACAG 855
QY 784 TTCTTTTTTGTAGACAGAGTCTCACTTTCATCACCTCAACCTCCGTTCCAGTCTCACTG 842
Db 856 GGAGTCTTGTCTGTTGTCCAGGCTGGAGTGTAGTGTGGGATCTCGGCTCCCTGCAAGA 915
QY 843 CTCTGCTCTCCGGGTTCAAAGYGAATCTCTGCTGAAGCTCTCTGAGTGTGGAAATTACA 902
Db 916 ATCTGCTCTCCAGATTACAGCAATTTTCCCACTCAGCCCCCTGGAGTAGTGAAGTACA 975
QY 903 AGGCTGACACACCATGCTTTGGCTAAT-----TTTTTGTATTTTGTAGAGATGGGGTTT 958
Db 976 GGCTCCCGCCACACCGCTGGCTAATTTTGTGTTTGTATTTTGTATTTTGTAGAGACGGGTTT 1035
```

Qy	959	ACCAATGTTGCCAGCGTGGTCTCAAACTCTGAGCTCAAGTGATCTGCGTGCCTCAGTCT	1018
Db	1036	ACTGTTAGGTAGGATGCTCGATCTCTGAGCTCA--TGATCCATCCGCTTGGCT	1093
Qy	1019	CCAAAGTCTGGAATTAATAGCGCTGAGTCACTGCTGCGGCCGATTAAGTCTATTTTC	1078
Db	1094	CTCAAAGTCTGGGATTAACAGGCATAGAGCCGACCGACCTGGCTG-----TTTT	1141
Qy	1079	TTTATTGCTATATCCAGATCTAGAGCAGTGTCTGACATATAGTAGTGCTCAATAAAT	1138
Db	1142	TTCAATTTCTATATGATTTGTTCTGTGGAATGTGAGTCACTTAGCGGTGTTTTACATG	1201
Qy	1139	AATTGATGAATGCACAGCCTAGATATAAATCTTTCTTTTCTTTTAAAAAATCTTGA	1198
Db	1202	TTTATAAGCTGAATAGTTCTGTATGTGAGTCAAGCACCCTAGTGTGAGTCACTACT	1261
Qy	1199	CAACTTTCAGAAATGAATACAACTCTTGCAATTTCTGTTTTTCACTTATCACCTGTTATGA	1258
Db	1262	CTCCTGCTGTGCATAGATGGGGGCTGTCCACACCCAAATCTCCTGGGTTCCCGCTGTCC	1321
Qy	1259	CTTTTTCATATGCTCAAACTTTATTTGTTACTGTTTTTTTCAITGTTACTATTTTAGTC	1318
Db	1322	GCACCTGAAGTACCCACAGATGTCTGTCACTGCCCTCCATCCACCTTCTTGTGGT	1381
Qy	1319	ACTGAATAATATGGCTTAATTTGCTTATATACATCTCTGCTGCCACTTTAGAGGCAAT	1378
Db	1382	TCTCATGTTTCTGCTC-----CCTAGAGACCAATTTCTCTACAGCAGTAG-----	1428
Qy	1379	TTACAAATCTGATGAAGCTATGAACCTCTCCACAGAGAAATACACACACACACACA	1438
Db	1429	-----TTTCCAATAGAAATACAAATGCCAGACACATA	1459
Qy	1439	CTCACACACAGTTTTTTTAAATGTTTGCACATAAGACAAAGAACTCGATTAGAGGATG	1498
Db	1460	C-----GTCATTTTAGATTTTCTAGTAGTCACTTTAGAAAGTGAAGAGGCCA	1510
Qy	1499	TTTGTTCATATTAATTAATAAATACTAGTTGGGCACAGTACTCAAGCTGTGTAACACA	1558
Db	1511	GCTG-----CAGTGGCTCACACCTGTAATCCCA	1538
Qy	1559	GTACTTTGAAGTCCNAGTGGTGGATCACTTGAGGTGAGAAGTTCGAGACCCAGCTGG	1618
Db	1539	GCACTTTGAAGGCTGAGGTGGTGGATCAC--GAGGTGAGGAGATCGAGACCATCTGG	1596
Qy	1619	TCAATATGTGAACCCCTATCTCTAATAAATAC--AAAAATTAGTGGGTCTAGTGATG	1677
Db	1597	CCAAATGTGTGAACCCCTCTCTACTAATAAATAAATAAATAAATAAATAAATAAATA	1656
Qy	1678	CATGCTGTAGTCCCAAGCTACTCGGGAGGCTGAGGCAAGAGAAATGCTTGAACCTGGGAG	1737
Db	1657	GGCGCTGTAGTCCCAAGCTACTCGAGAGGCTGAGGCAAGAGAAATGCGTGAACCCGGTAG	1716
Qy	1738	GAGAGGTTGAGTGAAGCGAGATCCACCACTGCACTCAGGCTGGGCGGACACAGCGAG	1797
Db	1717	GCAGAGCTTGCAGTGAAGCGAGATAGCGCCACTGCACTCCAGCTGGGCGGACAGAGGAG	1776
Qy	1798	ACTCTATCTCAAAAAATAAATAAATAAATAAAGGATCGGAGAGA	1843
Db	1777	ACTCCATCTCAAAAAATAAATAAAGGTTACAAAGAAACAGGTGAAA	1822
RESULT 10			
BO958903			
LOCUS			
DEFINITION			
BO958903 946 bp mRNA linear EST 21-AUG-2002			
AGENCOURT_10035485 NIH_MGC_40 Homo sapiens cDNA clone IMAGE:6483214			
5', mRNA sequence.			
BO958903			
VERSION			
BO958903.1 GI:22374381			
KEYWORDS			
EST.			
SOURCE			
Homo sapiens (human)			
ORGANISM			
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;			

Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini; Hominidae; Homo.

1 (bases 1 to 946)

NIH-MGC <http://mgi.nci.nih.gov/>.

National Institutes of Health, Mammalian Gene Collection (MGC) Unpublished (1999)

Contact: Robert Strausberg, Ph.D.

Email: sgabbs@mail.nih.gov

Tissue Procurement: DCTD/DTP

cDNA Library Preparation: The Rubin Laboratory

cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)

DNA Sequencing by: Agencourt Bioscience Corporation

Clone distribution: MGC clone distribution information can be found through the I.M.A.G.E. Consortium/LLNL at: <http://image.llnl.gov>

Plate: LRCM2667 row: 9 column: 23

High quality sequence stop: 534.

Location/Qualifiers

1..946

source

organism="Homo sapiens"

mol_type="mRNA"

db_xref="taxon:9606"

clone="IMAGE:6483214"

tissue type="carcinoma, cell line"

lab host="DH10B (phage-resistant)"

clone lib="NIH_MGC_40"

note="Organ: prostate; Vector: pOTB7; Site 1: XhoI; Site 2: EcoRI; cDNA made by oligo-dT priming. Directionally cloned into EcoRI/XhoI sites using the following 5' adaptor: GGCACGAG(G). Library constructed by Ling Hong in the laboratory of Gerald M. Rubin (University of California, Berkeley) using ZAP-cDNA synthesis kit (Stratagene) and Superscript II RT (Life Technologies). Note: this is a NIH_MGC Library."

ORIGIN

Query Match 15.0%; Score 300.6; DB 5; Length 946;

Best Local Similarity 69.1%; Pred. No. 8.9e-18;

Matches 468; Conservative 1; Mismatches 197; Indels 11; Gaps 4;

Qy 396 AGTTTGGGTTTTTTTTTTTGTGTTTGTAGACAGGGTCTTGCTCTGTCAAGGATG 455

Db 24 AGNTTTTTTTTTTTTTTTTTTTTGTGACAGAGTCTTGCTCTGTCAAGGATG 83

Qy 456 AGCAGAGTGTGCAACCAATAGTCACTGACGCTCAACCTCTGAGCTCAAGGATGTC 515

Db 84 AGTGAGTGTG---CGATCGCTCACTGCAACCTCAGCTCTGAGTTAAGTGAATCTC 140

Qy 516 TGACCTCAGCTCCCAAGTAGCTGGGACTACGAGCGTGCACACCCAGCTGGCTAAATTA 575

Db 141 CTGCTCAGCTCTCCAGTAGTCTGGATTAAGTGTGCAACCACTGCCGCTAAATTT 200

Qy 576 AAAAAATTTTTTTGTAGAGACTGGGTCTTAATAGCTTGGCAGGCTTGTCTTAAATCTCT 635

Db 201 TTTTGTATTTTGTAGTGAGACAGGGTTTACCATTTGGCCAGGCTGGTCTGCAATCTCT 260

Qy 636 GGCCTCAAGCAATCTCTCTAAGTGTGATCCCAAGTGTGGATTAAGGATTAAGGATTAAG 695

Db 261 GACCTC--GTGATCGGCTGCTCAGCTCCCAAGTGTGGGATTAAGGATTAAGGATTAAG 318

Qy 696 CCATGTGGGCTACTTATTTCTTTACATTTCCATCTTTCCAAATGAATGAATGAATGAAT 755

Db 319 CCAGCTTGGGCCCCAGTCCAAATATTTAAGATTTGTTTCTTAGTGTCTTGAAGTTTG 378

Qy 755 AACAGGATTAATGCTATATTTTCTTTCTTTTGTGAGACAGAGTCTCACTCAATCAAC 815

Db 379 CACA-----AAATTTCTTTTGTAGATGAGTCTCACTCTGCTACCCAGGCTGGAGTGC 433

Qy 816 CTCACCTCGGTCACTGACCTGCAACCTCTGCTCCCGGTTCAAGYATTTCTCTCTGCC 875

Db 434 AGTGGCGTATCTTGGCTCACTGCAACCTCTGCTCCCTCTGCTCCCTCAAGCAATTTCTCCACC 493

Qy 876 TAAGCTCTCTGAGTAGTGAATTAACAGCGTGCACCACTGCTTGGCTTAATTTTTTGT 935

Db 494 TCAGCCTCCAGTAGCTGGGATTACAGACGTGTGCACCACTCTGGGTAA-TTTTTGC 552
 Qy 936 ATTTTTCAGAGATGGGTTTTTACCATGTTCGCCAGGCTGTCTCAAACTCTCTGACCTC 995
 Db 553 ATTTTTCAGAGAGAGGAGTTTACCATGTTCGCCAGGCTGTCTGAACTCTCTGACCTC 612
 Qy 996 AAGTATCTGCTGCTCTAGTCTCCCAAGTCTCGGAATATAGCGGTGAGTCACTGTGC 1055
 Db 613 AGGTATCTCTGCTGCTGCTCCCAAGTCTCGGAATATAGCGGTGAGTCACTGTGC 672
 Qy 1056 CTGGCGGATTACTGTCT 1072
 Db 673 TCAGCCGCNAATCTCT 689

RESULT 11
 B0706343/c
 LOCUS
 DEFINITION BQ706343 918 bp mRNA linear EST 16-JUL-2002
 AGENCOURT_8484983 NIH_MGC_113 Homo sapiens cDNA clone IMAGE:6301091
 5', mRNA sequence.

ACCESSION BQ706343
 VERSION BQ706343.1 GI:21845242
 KEYWORDS EST.
 SOURCE Homo sapiens (human)

ORGANISM
 Homo sapiens
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini;
 Homidae; Homo.

REFERENCE
 1 (bases 1 to 918)
 NIH-MGC <http://mgs.nci.nih.gov/>.
 National Institutes of Health, Mammalian Gene Collection (MGC)
 Unpublished (1999)

AUTHORS
 TITLE
 JOURNAL
 COMMENT
 Contact: Robert Strausberg, Ph.D.
 Email: cgapbs-remail.nih.gov
 Tissue Procurement: Dr. Mark Watson
 cDNA Library Preparation: Rubin Laboratory
 DNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
 DNA Sequencing by: Agencourt Bioscience Corporation
 Clone distribution: MGC clone distribution information can be
 found through the I.M.A.G.E. Consortium/LLNL at:
<http://image.llnl.gov>

Plate: LLCM2516 row: 9 column: 12
 High quality sequence stop: 632.

FEATURES
 source

1. 918
 Location/Qualifiers
 /organism="Homo sapiens"
 /mol_type="mRNA"
 /db_xref="taxon:9606"
 /clone="IMAGE:6301091"
 /lab_host="DH10B (phage-resistant)"
 /clone_lib="NIH MGC_113"
 /note="Organ: spleen; Vector: pOTB7; Site 1: XhoI; Site 2:
 EcoRI; cDNA made by oligo-dT priming. Directionally cloned
 into EcoRI/XhoI sites using the following 5' adaptor:
 GGACACAG(G). Library constructed by Ling Hong in the
 laboratory of Gerald M. Rubin (University of California,
 Berkeley) using ZAP-cDNA synthesis kit (Stratagene) and
 Superscript II RT (Life Technologies). Note: this is a
 NIH_MGC Library."

ORIGIN

Query Match 14.9%; Score 298; DB 5; Length 918;
 Best Local Similarity 68.6%; Pred. No. 1.5e-17;
 Matches 456; Conservative 1; Mismatches 197; Indels 11; Gaps 3;
 Qy 398 TTTGGGTTTTTTTGTGTTTGTAGAGACAGGCTTGTCTGTCAACCCAGGCAATGAG 457
 Db 666 TTTTGGATTATGATTATTATTTTGAAGCCAGGCTCTGCTCTGTCAACCCAGGCGGAG 607
 Qy 458 CACAGTGTGAACATAGTCACTGCGGCTCAACCTCTGTAGCTCAAGGATCTGCTG 517
 Db 606 TGCAGTGTGAATCATGTGGTCACTCCAGGCTTTATCTCTCTGGGCCCAAGTATCTTCCA 547

Qy 518 ACCTCAGCCTCCCAAGTAGCTGGGACTACGAGCGTGCACCAACCGCTGGCTA----- 571
 Db 546 ACCTCAGCCTCCCAAGTAGCTGGGATTATAAGTGGTGTCTAGGATGTCTGGCTAATTTT 487
 Qy 572 -ATTAAAAAATTTTTTTGTAGAGACTGGGTCTTACTAGTGTGGCCAGGCTTGTCTTAAA 630
 Db 486 TATTATATAATTTTTTTGTAGAGACAGGCTCTGGCTTTGTGGCCAGGCTGGTCTCAA 427
 Qy 631 CTCCTGGCTTCAAGCAATCCCTACCTTGGCATCCCAAGTGTGGGATTTACAGGCGTG 690
 Db 426 TTCCTGGACTCAGCGGATCTCTGCTTTCAGCCTCCCAAGTGTGGGATTTACAGCGCTG 367
 Qy 691 AGCC---ACCATGTGGCGCTACTTATTCTTTTACATTCATCTTTTCCAAATAGAATGTAAG 747
 Db 366 AACCTTTCAACTTTTAAATGCTTTTTCAGTATTTTCCATCCCTTGATTGACATTTGGGTTT 307
 Qy 748 ATCCACAGAACAGGATTTACTGCTATTTTCTCTCTTTTGTGAGACAGAGTCTCAC 807
 Db 306 TAAATTTATTTTGTGTGTTGTTCTTTTGTAGACAGAGTCTTGTCTGTCAACCGGCT 247
 Qy 808 TTCTATCACTCAACCTCGGTTTCACTGCTCACTGCAACCTCTGCTCCCGGTTCAAGYATT 867
 Db 246 TAGAATGCGGGGTGTGATCTCGCTCACTGCAACCTCTGCTCTCTCGTTCAAGCGATT 187
 Qy 868 CTCCTGCTAAGCCTCTCTGAGTAGCTGGAATTACAAGCTGCACCACTCTTGGCTTAA 927
 Db 186 CTTGTGCTCAGCCTCTCTGAGTAGCTGGGATTACAGGCGGCTGCATGATGCCCACTAA 127
 Qy 928 TTTTCTGATTTTTPAGCAGAGATGGGTTTATACATGTGCCAGGCTGGTCTCAAACTC 987
 Db 126 TTTTCTGATTTTTPAGCAGAGATGTGTTTCCATGCTGGCCAGGCTGGTCTCAAACTC 67
 Qy 988 CTGACCTCAAGTATCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 1046
 Db 66 CTGACCTCAAGTATCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 7
 Qy 1047 TCACCT 1051
 Db 6 CCACT 2

RESULT 12
 CX780936
 LOCUS
 DEFINITION HESC3_2.B04.b1.A036 NIH_MGC_260 Homo sapiens cDNA clone
 IMAGE:30928759 3', mRNA sequence.
 ACCESSION CX780936
 VERSION CX780936.1 GI:58297726
 KEYWORDS EST.
 SOURCE Homo sapiens (human)
 ORGANISM
 Homo sapiens
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini;
 Homidae; Homo.

REFERENCE
 1 (bases 1 to 736)
 NIH-MGC <http://mgs.nci.nih.gov/>.
 National Institutes of Health, Mammalian Gene Collection (MGC)
 Unpublished (1999)

AUTHORS
 TITLE
 JOURNAL
 COMMENT
 Other ESTs: HESC3_2.B04.g1.A036
 Contact: Daniela S. Gerhard, Ph.D.
 Office of Cancer Genomics
 National Cancer Institute / NIH
 Bldg. 31 Rm10A07 Bethesda, MD 20892
 Email: cgapbs-remail.nih.gov
 Tissue Procurement: BresaGen, Inc.
 cDNA Library Preparation: Express Genomics, Inc.
 cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
 DNA Sequencing by: Laboratory for Genomics and Bioinformatics,
 University of Georgia
 Clone distribution: MGC clone distribution information can be
 found through the I.M.A.G.E. Consortium/LLNL at:
<http://image.llnl.gov>

Plate: NDA1172 row: c column: 08
Seq primer: M13-21 (TGTAACACGACGGCCAGT)
High quality sequence stop: 711
POLYA=Yes

Location/Qualifiers
1. .736
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/clone="IMAGE:30928759"
/sex="male"
/tissue_type="human embryonic stem cells"
/cell_type="human embryonic stem cells"
/cell_lines="BG01"
/lab_host="NIH MGC 260"
/clone_lib="NIH MGC 260"
/note="Vector: pExpress-1; Site 1: NotI; Site 2: EcoRV;
RNA obtained from human embryonic stem cells isolated from
the inner cell mass of blastocyst stage embryos. Cell line
id and NIH Registry designation is BG01. Positive for
SSEA3, SSEA4, Tra 1-60, Tra 1-81, CD9, Alk Phos, Oct4 and
Nanog expression; negative for SSEA1 expression. Passage
number 21. CDNA primed using oligo-dT primer:
5'-pGACTAGTCTAGATCGAGCGCGCC(T)25-3' and cloned into
the EcoRV/NotI sites of pExpress-1. This primary library
is non-normalized (normalized by Express Genomics
NIH MGC 261). It was constructed by Express Genomics
(Frederick, MD). Sequence ends have been trimmed to
exclude vector and regions below Phred quality 16. Note:
this is a Mammalian Gene Collection library."

ORIGIN
Query Match 14.7%; Score 293.6; DB 8; Length 736;
Best Local Similarity 66.9%; Pred. No. 4.3e-17;
Matches 464; Conservative 1; Mismatches 220; Indels 9; Gaps 3;

QY 399 TTGGGTTTTTTTTTGTGTTTGTAGACACAGGGTCTGCTGTCAACCAGGCAATGAGC 458
DB 1 TTTTITTTTTTTTTTTTTTTTTTGTGATGGAGTCTACTCTCTGCGCAGGCTGAAGT 60

QY 459 ACAGTGGTGCACCATAGTCACTGAGCTCAACTCTCTGAGTCAAGGATCTGCTGA 518
DB 61 GCAGTGGCGGGATCTCGGCTCACTGCAACTCTGCTCTGGGTCAAGCACTCTCTTG 120

QY 519 CCTCAGCCTCCCAAGTAGCTGGAGTACGAGGTGCACCAACGCTGGCTAAATTA 578
DB 121 CTTTACGCCCCCTGAGTGGATTTAGGACCAACCAACGCTGGATCTT-TT 178

QY 579 AAATTTTTTTGTAGACAGTGGTCTTACTACGTTGGCCAGGCTTGTCTTAAACTCTGCG 638
DB 179 TTGTATTTTGTAGTAAATATGGGTTTGTCTATATTGGCAGGCTGTCTCAAACTCTGAC 238

QY 639 TTCAAGCAATCTCTCACTTGGCATCCCAAGTGTGGATTTACAGGGGTGAGCCACCA 698
DB 239 CTGAGTGTATCCACCTGCCTCAAGCTGCGCAAGTGTGGATTTACAGGGATGAACACCA 298

QY 699 TGTGCGGCTACTTATTTCTTTTACATTCATCTTCCATTTCCAATAGATGTAAGATCCACAGA 758
DB 299 TGCCAGGCTCTTATTTCTTTTATATATATATCAAGTATTTATTTATTTATTTA 358

QY 759 AGGGATTTACTGCGCTATTTCTCTTTCTTTTGTAGACAGTCTCACTTCAATCACTC 818
DB 359 TTTATTTATTTTGTAGATGGATTTTGGCCCTTGTGGCCAAAGTGGAGTGAATGGCAC 418

QY 819 AACCTCGTTCAGTCACTGCAACCTCTGCTCCCGGTTCAAGYGAATTTCTCTGCTTAA 878
DB 419 GATC-----TCAGTCTCACTGCAACCTCCACCTCCCGGGTTCAAGTGAATTTCTCTGCTCA 473

QY 879 GCCTCCCTGAGTGTGAATTTACAGGCTGCACCAACCATGCTTGGCTAAATTTTGTAT 938
DB 474 GCCTCCCTGAGTGTGGATTTACAGGCAATGCCATCATGCTGCTTAA--TTTGTCA 531

QY 939 TTTAGCAGAGATGGGGTTTTACCATGTTGGCCAGGCTGGTCTCAAACTCTGACCTCAAG 998

Db 532 TTTAGTAGACACAGAGTTTCTCCATGTTGTGTCAGGCTGGTCTTGAACCTCTGACCTCAGG 591

QY 999 TGATCTGCGCTGCTCAGTCTCCCAAGTGTCTGAATATAGGCGTGAGTCACTGTGCGCTG 1058

Db 592 TGATCTGCGCTGCTCAGGCTCCCAAGTGTCTGGGATTTATAGGATGAGCCACCGTGCCCA 651

QY 1059 GCCGATTACTGTCTATTTCTTTTATTGCTATATC 1092

Db 652 GCCATTATCAAGATTATTTATTTATTTCTTCTTCTC 685

RESULT 13

CR613629 3092 bp mRNA linear HTC 21-JUL-2004

LOCUS full-length cDNA clone CSODI039YN19 of Placenta Cot 25-normalized of Homo sapiens (human).

DEFINITION

CR613629

ACCESSION

VERSION CR613629.1 GI:50494436

KEYWORDS HTC; CNSLT_cDNA

SOURCE Homo sapiens (human)

ORGANISM Homo sapiens

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini; Hominidae; Homo.

1 (bases 1 to 3092)

Li,W.B., Gruber,C., Jessee,J. and Polayes,D.

Full-length cDNA libraries and normalization

Unpublished

JOURNAL

REMARK

Contact : Feng Liang Email : fliang@lifetech.com URL : http://fulllength.invitrogen.com/ Invitrogen Corporation 1600 Faraday Avenue

2 (bases 1 to 3092)

Genoscope.

Direct Submission

TITLE

JOURNAL

COMMENT

1st strand cDNA was primed with a NotI-oligo(dT) primer. Five prime end enriched, double-strand cDNA was digested with Not I and cloned into the Not I and EcoR V sites of the pCMVSPORT 6 vector. Library was normalized. Library was constructed by Life Technologies, a division of Invitrogen.

FEATURES

source

1..3092

/organism="Homo sapiens"

/mol_type="mRNA"

/db_xref="taxon:9606"

/clone="CSODI039YN19"

/tissue_type="Placenta Cot 25-normalized"

/plasmid="pCMVSPORT_6"

ORIGIN

Query Match 14.5%; Score 290.6; DB 4; Length 3092;
Best Local Similarity 69.7%; Pred. No. 3.1e-17;
Matches 466; Conservative 1; Mismatches 190; Indels 12; Gaps 5;

QY 394 ACAGTTTGGGTTTTTTTTTGTGTTTGTAGACAGGCTTGTCTCTGACCCAGGCA 453

DB 1750 AAATTTTTTTTTTTTTTTTTTTTTTTTGTAGACAGAGTCTTGTCTGTCAACCCAGGCT 1809

QY 454 TGACACAGTGTGTGCAACCATAGGTCACTGACGCTCAACCTCTGAGCTCAAGGATCT 513

DB 1810 GGAGTGCAGTGGTG---CGATCGCTCACTGCAACCTCAACCTCTCTGGATTTAAGTATTC 1866

QY 514 GCTGACCTCAGCTCCCAAGTAGCTGGAGTACGAGCGTGCACACACCTGGCTAAT 573

DB 1867 TCTGCTCAGCTCCCAAGTAGCTGGGATTTACAGGTGTGCACCAACCATGCCCGCTAAT 1926

QY 574 TAAAAAAATTTTTTTGTAGACAGTGGGTCTTACTAGTTGGCCAGGCTTGTCTTAACTC 633

DB 1927 TTTTTCGTA-TTTTTAGTGAGACAGGCTTTCCACCATGTTGGCCAGGCTGGTCTCGAATC 1985

1 (bases 1 to 769)
NIH-MGC <http://mgc.nci.nih.gov/>.
National Institutes of Health, Mammalian Gene Collection (MGC)
Unpublished (1999)
Contact: Robert Strausberg, Ph.D.
Email: rgabs-r@mail.nih.gov
Tissue Procurement: Dr. Mark Watson
cDNA Library Preparation: Rubin Laboratory
cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
DNA sequencing by: Agencourt Bioscience Corporation
Clone distribution: MGC clone distribution information can be
found through the I.M.A.G.E. Consortium/LLNL at:
<http://image.llnl.gov>
Plate: LLCM2476 row: d column: 21
High quality sequence stop: 549.
Location/Qualifiers
1. 769
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/clone="IMAGE:6282212"
/lab_host="DH10B (phage-resistant)"
/clone_lib="NIH_MGC_113"
/note="Organ: spleen; Vector: pOTB7; Site_1: XhoI; Site_2:
EcoRI; cDNA made by oligo-dT priming. Directionally cloned
into EcoRI/XhoI sites using the following 5' adaptor:
GGCAGCAG(G). Library constructed by Ling Hong in the
laboratory of Gerald M. Rubin (University of California,
Berkeley) using ZAP-cDNA synthesis kit (Stratagene) and
Superscript II RT (Life Technologies). Note: this is a
NIH MGC Library."

ORIGIN

Query Match	14.4%	Score 287.6	DB 5	Length 769
Best Local Similarity	67.0%	Prod. No. 1.4e-16		
Matches 453	Conservative 1	Mismatches 215	Indels 7	Gaps 3
QY	404	TTTTTTTTTTGTTGTTTATGAGACAGGGTCTGTGCTCTGTCCACCCAGGCATGAGACAGT	463	
Db	768	TTTTTTTTTTTTTTTTTTATGAGACAGAGTCTCACTGTCTCCCGAGGCTGAGTGCAGT	709	
QY	464	GGTGCACCATATGAGTCACTGAGCGCTCAACCTCTCTGAGCTCAAGGGATCTGCTGACCTCA	523	
Db	708	GGGGTGATCTCGGCTCACTGCAACCTCTGCTCTCTGGGTTCAAGCGATTTCTCTCGCTCA	649	
QY	524	GCCTCCCAAGTAGCTGGGACATGAGCGGTGACACACACGCGCTGGCTAATTAATAAAATTT	583	
Db	648	GCCTCCGAGTAGCTGGGACATATAGCGCCCGCCCACTGTGCGGGCTAATTTTATATA---T	592	
QY	584	TTTTTGTAGAGACTGGGTCTTACTACGTTGGCCAGGCTTGCTTTAAATCTCTGGCTTCAA	643	
Db	591	TTTTTAGTAGACGGGGTTTCCCAATGTGGCCAGGCTGGTCTCGAATTTCTTAACCTCAG	532	
QY	644	GCAATTCCTCTACCTTGGCATCCCAAGTGTCTGGGATTAACAGGGGTGAGCCACCATGTGC	703	
Db	531	GTGATCCGGCTGGCTCAGCCTCCCAACATGCTGGGATTAACAGTGTGAGCCACCATGCC	472	
QY	704	GGCTACTTAATTTCTTTTATATCCATCTTTTCCAAATAGAATGTAAAGATCCACAGGA	763	
Db	471	GGCTGTGTATACACTTTTTTTGTGGTGTCTACAGGCTCTCTGAGGCTCTGTCTCAATTTTGT	412	
QY	764	TTACTGCCTATTTTTCTTCTCTTTTGTGAGACAGAGTCTCACATCTCATCACTCAACCT	823	
Db	411	TTGTTTTTATTTTTTTTGAGACAGAGTCTCTTTCTGTCACTAGACTGGAGTGCAGTGGTGC	352	
QY	824	CGGTTTCAGCTCACTGCAACCTCTGCTCTCCGGGTTCAAGYGATTTCTCTGCCATAAGCCTC	883	
Db	351	AATCTCGGCTCACTGCAACCTCTGCTTCAGAGTTCCAGCGATTTCTTGTCTCAGCCTC	292	
QY	884	CTGAGTAGCTGGAATTTACAAGCGTGTGACCAACATGCTTGGCTTAA---TTTTTGTATTTTT	941	
Db	291	CCAAGTAGCTGGGATTAACAGGCATGCCACATGCCCCAATTAATTTTTTTTTTGTATTTTT	232	

GenCore version 5.1.7
Copyright (c) 1993 - 2006 Bioceleration Ltd.

OM nucleic - protein search, using frame_plus_n2p model

Run on: February 8, 2006, 09:18:47 ; Search time 65.2558 Seconds
(without alignments)
2694.617 Million cell updates/sec

Title: US-10-607-806-1_COPY_7200_9200

Perfect score: 3524

Sequence: 1 ctatctcagctgtccctccc.....ckctacataaatatctatta 2001

Scoring table:

BLASUM62
Xgapop 10.0 , Xgapext 0.5
Ygapop 10.0 , Ygapext 0.5
Fgapop 6.0 , Fgapext 7.0
Delop 6.0 , Delext 7.0

Searched: 2443163 seqs, 439378781 residues

Total number of hits satisfying chosen parameters: 4886326

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Command line parameters:

-MODEL=frame+ n2p.model -DEV=xlp
-O=/abs/ABSSWEB.spool/US10607806/runat_08022006_091817_26823/app_query.fasta_1
-DB=A_Geneseq -QFMT=fastan -SUFFIX=n2p.rag -MINMATCH=0.1 -LOOPEXT=0 -LIST=45
-UNITG=bits -START=1 -END=1 -MATRIX=blsum62 -TRANS=human40.cdi -MODE=LOCAL
-DOCALIGN=200 -THR SCORE=pct -THR MAX=100 -THR MIN=0 -ALIGN=15 -MODE=LOCAL
-OUTFMT=ptc -NORM=ext -HEAPSIZE=500 -MINLEN=0 -MAXLEN=2000000000 -HOST=abs02p
-USER=US10607806 @CGN 1.1.605 @runat_08022006_091817_26823 -NCPU=6 -ICPU=3
-NO MMAP -NEG SCORES=0 -WAIT -DSPLOCK=100 -LONGLOG -DEV TIMEOUT=120
-WARN TIMEOUT=30 -THRAADS=1 -XGAPOP=10 -XGAPEXT=0.5 -FGAPOP=6 -FGAPEXT=7
-YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database :

- 1: Geneseq21.*
- 2: Geneseq1980s.*
- 3: Geneseq1990s.*
- 4: Geneseq2000s.*
- 5: Geneseq2001s.*
- 6: Geneseq2002s.*
- 7: Geneseq2003as.*
- 8: Geneseq2003bs.*
- 9: Geneseq2004s.*
- 9: Geneseq2005s.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
C 1	459	13.0	7067	6	Aae35265 Human P45
2	449.5	12.8	7052	6	Aae35266 Human P45
C 3	444	12.6	7052	6	Aae35266 Human P45
4	438.5	12.4	215	4	Aau31513 Novel hum
C 5	432	12.3	7036	6	Aae35264 Human P45
6	427.5	12.1	217	7	Adc87151 Human GPC
7	420.5	11.9	7067	6	Aae35265 Human P45
8	416	11.8	673	6	Ada84009 Human POM
9	413	11.7	185	4	Aao06014 Human pol

10	405	11.5	232	5	ADK36978	Adk36978	Novel hum
11	400	11.4	302	5	ADK36936	Adk36936	Novel hum
12	400	11.4	361	5	ABG68738	Abg68738	Human pro
13	400	11.4	449	5	ABJ01057	Abj01057	Ovary cel
14	397.5	11.3	304	4	AAU03091	Aau03091	Novel hum
15	397.5	11.3	375	3	AAU03399	Aau03399	Neuron-as
16	397.5	11.3	375	5	ABB81538	Abb81538	Neural th
17	397.5	11.3	375	5	AAE29142	Aae29142	AD7C-NTP
18	397.5	11.3	375	6	ABG63268	Abg63268	AD7C-NTP
19	397.5	11.3	375	6	ADA84017	Ada84017	Human POM
20	397.5	11.3	375	6	ABU03024	Abu03024	Human neu
21	397.5	11.3	375	6	ABB99774	Abb99774	Amino aci
22	397.5	11.3	375	6	ABJ19445	Abj19445	AD7C-neur
23	397.5	11.3	375	7	ADB37642	Adb37642	Human neu
24	397.5	11.3	375	8	ADR14409	Adr14409	Human NF-
25	397.5	11.3	375	9	AEA30020	Aea30020	Human neu
26	396.5	11.3	239	8	ADG65417	Adg65417	Novel hum
C 27	391.5	11.1	673	6	ADA84009	Ada84009	Human POM
28	390	11.1	382	4	AAU32610	Aau32610	Novel hum
29	390	11.1	382	4	AAU31818	Aau31818	Novel hum
30	390	11.1	382	4	AAU32707	Aau32707	Novel hum
31	388	11.0	7036	6	AAE35264	Aae35264	Human P45
32	387.5	11.0	396	4	AAU30455	Aau30455	Novel hum
33	386	11.0	179	8	ADR09211	Adr09211	Human pro
34	384.5	10.9	158	7	ADB64893	Adb64893	Human pro
35	384	10.9	324	4	AAU29573	Aau29573	Novel hum
36	384	10.9	324	4	AAU29579	Aau29579	Novel hum
37	384	10.9	397	2	AAE95913	Aae95913	Neural th
38	383.5	10.9	257	4	AAO10294	Aao10294	Human pol
39	381	10.8	241	8	ADQ66812	Adq66812	Novel hum
40	381	10.8	527	7	ADD27133	Add27133	Human adi
41	380.5	10.8	175	4	AAU31857	Aau31857	Novel hum
42	380.5	10.8	250	4	AAU31823	Aau31823	Novel hum
43	380.5	10.8	250	7	ADE09441	Ade09441	Novel pro
44	377	10.7	215	4	ABG07921	Abg07921	Novel hum
45	376	10.7	189	7	ADB65042	Adb65042	Human pro

ALIGNMENTS

RESULT 1

AAB35265

ID AAE35265 standard; protein; 7067 AA.

AC AAE35265;

DT 28-MAY-2003 (first entry)

DE Human P45ORAI-3 protein #2.

KW Human; retinoic acid; vitamin A; cytochrome P450; actinic keratosis;

KW P45ORAI; cancer; psoriasis; acne; ichthyosis; gene therapy; vaccine;

OS Homo sapiens.

PN WO200295034-A2.

PD 28-NOV-2002.

PF 23-MAY-2002; 2002WO-CA000758.

PR 23-MAY-2001; 2001US-0292531P.

XX (CYTO-) CYTOCHROMA INC.

XX Wisniewski J, Petkovich PM, Ramshaw H;

XX WPI; 2003-112152/10.

XX N-PSDB; AAD53861.

XX New isolated P45ORAI-3 nucleic acid molecule and polypeptide, useful for
diagnosing, preventing, or treating disorders with aberrant expression or

PT activity of the P450RAI-3, such as cancer, actinic keratosis, acne and
PT psoriasis.
XX
XX Example 1; Fig 1; 231pp; English.
XX
CC The invention relates to retinoic acid metabolising cytochrome P450,
CC P450RAI polypeptides and polynucleotides. Methods and compositions of the
CC invention are useful for diagnosing, preventing, ameliorating and/or
CC treating disorders associated with the aberrant expression or activity of
CC the P450RAI such as diseases related to vitamin A and retinoic acid
CC metabolism, e.g. cancer, actinic keratosis, psoriasis, acne and
CC ichthyosis. The invention is useful in gene therapy and as vaccine. The
CC present sequence is human P450RAI protein
XX
SQ Sequence 7067 AA;

Alignment Scores:
Pred. No.: 2,6e-43 Length: 7067
Score: 459.00 Matches: 196
Percent Similarity: 40.9% Conservatives: 59
Best Local Similarity: 31.4% Mismatches: 186
Query Match: 13.0% Indels: 183
DB: Gaps: 25

US-10-607-806-1_COPY_7200_9200 (1-2001) x AAE35265 (1-7067)

QY 1977 AAGTTGCAFTATTTTGGCTTCTTACTTTTATAACCAATAAATAAGATTACAGT 1918
DB 77 LysValTrpTyrIleHisThrIleTyrTyrLeuSerIleLysArgAsnGluValIle 96
QY 1917 TAAATGCATTTAACTTTATACATATTAATTACGTATCTCTGCTTACCTTCAGA--- 1861
DB 97 MethHisAlaThrThrLeuLysIle-----LeuCysLeuSerGluArgSer 112
QY 1860 -----ATCTATTAGTTTGTCTCTCCGATCCTTTATTAT 1822
DB 113 GlnLeuGlnLysThrArgValPheLeuPheCysLeuPhe-----CysPheVal 128
QY 1821 TTATTATTATTGAGATAGAGTCTCGCTGTGTGCGCCAGGCTGAGTGCAGTGTGGG 1762
DB 129 LeuPheCysPheLysArgGlnGlyLeuAlaLeuLeuProIleLeuGlu-Tyr-----Se 146
QY 1761 ATCTCGGCTCAGTGCAGCTCTGCTCCAGGTTCCAGCAATCTCTGCTCAGCTCC 1702
DB 146 rSerLeuLeuThrAlaAlaSerThrPheTrpThrGlnAlaSerLeuProGlnProPr 166
QY 1701 CGAGTAGCTGGGACTACAGCATCATCACTACCCAGCTAATTTTGTATTTTAGTA 1642
DB 166 Olys-----LeuAlaLeuGlnValCysThrThrThrAlaSer---PheLeu-IleValCysE 184
QY 1641 GAGATAGGTTTCCACATATTGACAGGCTGTCTCGAACTTCTCACTCAAGTGATCCA 1582
DB 184 rGlnArgValSerProCysCysProGlyTrpSerGlnThrProSerLeuLys----- 201
QY 1581 CCCACCTTGACTTCCAAAGTACTGTGGTTACAGGCTTGAGTCAGTCTGCCCACTGAGT 1522
DB 202 -----ProTrpProL 205
QY 1521 TATTTTAAATATATGACAAACATCCCTAATGACAGGTTTCTTCTTAGTTGCAAAC 1462
DB 205 eulysValLeuGlyLeuGlnGlyAlaThrMetValGly----- 217
QY 1461 ATTAATAAATACTGTGTGAGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 1409
DB 218 -----GlnLysThrThrCysTyrLeuIleProIle-----TrpGlyValIleA 232
QY 1408 -----GAGGTTCTAGCTTTTCATCAGATTGTGTAATTTGGC 1372
DB 232 laAsnGlyTyrGlyValSerPheArgGlySerLysIleAspAspGly-CysThr----- 249
QY 1371 CTTCCTAAAGTGGAGGAGGATGATTAACCAATTAAGCCATATTATTCAGTGACTAAA 1312
DB 250 IleLeuIleGluSerLysProLeuIleCysAlaLeuMetSerLeuTyrGlyThrSerGln 269

QY 1311 ATAGTAACAATGAARAAAC-----AGTAACAATAAAGTTTGTAGGCAATATGA 1264
DB 270 SerCysLeuLys-LysAsnPheSerTyrIleThrArgAsnArgGlyLeu----- 285
QY 1263 AAAAGTCATAACAAGGTGATAAGTGAATAAAGCAAGATGCAAGATTGTATTATTCTGCAA 1204
DB 285 ----- 285
QY 1203 AGTTGTCAAGATTCTTTTAAAAAAGAAAGTTTATATCTAGCTGT----- 1152
DB 286 -----IleIleLeuLysLeuLysGlnLysMetValGlyLeuProLysLy 302
QY 1151 -----GCATTCACTAATTATTTATTGAGC-----ACCTACTATATGTC 1114
DB 302 sArgAlaGlyIleLysGlyLysPheSerTyrValIleAlaValLysThrGlnAsnLysAs 322
QY 1113 AGACACTGCTCTAGATCTGGGGATATAGCAATAAAGAAATATAGACAGTAATCGGCCAGGC 1054
DB 322 nLysProIleArgAspSerGluIleIleThrArgThrLysArgGlyValLysGlyGlnTh 342
QY 1053 ACAGTGACTACGCTCATATATTCAGCACTTTGGAGACTGAGGAGGAGCAGATCACTTGA 994
DB 342 rGlnTrpLeuThrProValIleProAlaPheTrpGluAlaGluValGlyGlySerLeuGl 362
QY 993 GGTGAGGATTGTAGACCGCTGGCAACATGTTAAACCCCATCTCTCTAAAAATAC 934
DB 362 uAlaArgSerLeuArgProAlaTrpAlaThrTrp---AsnProValSerThrLysAsnTh 381
QY 933 AAAAAATAGCAAGCATGTGTGTGCACGCTTGTAAATTCACAGTACTCAGGAGGCTTAGG 874
DB 381 r-LysIleSerArgAlaTrpTrpTyrValProValProAlaAlaTrpGluAlaGluA 401
QY 873 CAGGAGATCTCTGAACCCGGAGGAGAGGTTGTCAGTGAGCTGAACGAGGTTGAGGT 814
DB 401 laGlnGluLeuLeuPro----- 407
QY 813 GATGAGTGACACTCTGTCTCAAAAAAGAAAGAAATAGGCAGTAATCCCTGTTCT 754
DB 407 ----- 407
QY 753 GTGGATCTTACATTTCTATTGGAAGATGGAATGTAAGAAATAAGTAGCCGACATGTGT 694
DB 408 -----GluAlaArgCysGlyG 413
QY 693 GCTACCCCTGTAAATCCAGCACTTTGGATGTCGAAGTAGGAGGATGCTTGAAGCCAG 634
DB 413 LysArgLeu---SerGlnHisPheGlyArgProArgTrpValAsp---HisGluValAr 431
QY 633 GAGTTTAAAGCAAGCC-----TGSCCAACGTAGTAAGACCCAGTCTCTACAAAAAAT 580
DB 431 gSerLeuArgProAlaProThrTrp-----AsnProIleSer---AsnLysTy 446
QY 579 TTTTAAATTAGCCAGGCGT 520
DB 446 rLysLysLeuAlaGlyHisGlyGlyVal---CysLeuCysGlnLeuLeuArgLeuAr 465
QY 519 GTCAGCAGATCCCTTGTAGCTCAGGAGTTGAGCTGCGATGACCTTGTGTCACCACCTG 460
DB 465 gGlnGluAsnCysLeuAsnLeuGlyGlyGlyCysSerGluProArgSerArgHisCy 485
QY 459 TCTCATGCTGCTGTGTCAGCAGACACCTGTCTCTAAACCAACCAACCAACCAACCA 400
DB 485 sThrProThrTrpAlaThrLysArgAspSerValSerLysLysLysLysLysLysG 505
QY 399 AACTGTCCACTTACAAAGCTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 340
DB 505 uLeuLeuGluSerGlyArgArgArgLeuGlnAlaGluIleAla-----ProLe 521
QY 339 GTTTTCTTATCTGTGAAGTGAGGATAATTGTACTATCCCATAGTGTGTTCTTGAGGATT 280
DB 521 uHisSerSerLeuGlyAspArgAlaArgLeuCys-----Le 533


```
Db      210  lyGluPro-----ProTTPalaArgA 217
QY      1352  GGATGATATAAGCAAAATAAGCCATAT-----TATP 1323
Db      217  rgProHis---ValIlePheProPheGlyGluLeuMetAlaThrGlyPheLeuSerG 236
QY      1322  CAGTGACTAAATAGTAAACAAAGTAAACAAACAGTAAACAAATAAAGGTTTGAGGCAATATGAA 1263
Db      236  lyAlaLeuLysLeuIleGluMetValAlaGln-----PheCysG 249
QY      1262  AAAGTCATACAAAGGTGATAAGTGAAGAAAGCAGAAATGC----- 1225
Db      249  luSerGlnAsnHisPheValProPheLysValAsnCysMetValHisLeuLysAlaValL 269
QY      1224  --AAGATTCTATTTATCTGCAAGTTGTCAAGATTGTTT-----AAAAAAA 1179
Db      269  ysLysIlePheLysValIleGlnGluValLeuTyrSerTyrLysAsnLysAsnA 289
QY      1178  GA-----AAAGAAAGTTTATATCTAGGCTGTGCATTCATCAAT 1140
Db      289  rgGlnLysTrpAspSerTyrProArgGluLeuValSerLysValAsnSerHisMetL 309
QY      1139  TATTTATTGAGCACCTACTATATGTCTAGTCTAGATCTCGGGATATAGCAATAA 1080
Db      309  eu-----AsnAsnLysLeuLysIleArgLysIleSerLeuLeuGlyIleValArgSerL 327
QY      1079  AGAAATAGACAGTAA---TCGGCCAGGCACAGTCACTCAGCCCTATAATTCACGACT 1024
Db      327  euGluLeuLysGluGluLeuLysAlaArgArgSerGlySerHisLeu---SerGlnHisP 346
QY      1023  TTGGGAGACTGAGGAGGAGATCACTTGAGTCAAGGAGTTTGAGACCAAGCTGGGCAAC 964
Db      346  heGlyArgProArgTrpValAspHisLeuArgProGlyVal---AspGlnProGlyGlnH 365
QY      963  ATGTTAAACCCCTCTCTGCTAAATAACAAAAATTAGCCAAAGCATGCTGGTGCACGC 904
Db      365  isGlyGluThrProSerLeuGlnLysThrGln-LysLeuAlaGlyArgGlyGlyMetCys 384
QY      903  TTGTAATTCAGTACTCAGGAGGCTTAGCGAGAGATCRCTTCAAC-----CCGGGA 850
Db      385  Leu---SerGlnLeuLeuGlyArgLeuArgHisLysAsnCysLeuAsnLeuArgProGly 403
QY      849  GGCAGAGTTGAGTGAGTGAACGGAGGTTGAGTGATGAAGTGAGACTCTGTCTCAA 790
Db      404  Ala----- 404
QY      789  AAAGAAAGGAAGAAATAGGCAGTATCCCTGTTCTGTGGATCTTACATTTATTTGAAA 730
Db      404  ----- 404
QY      729  GATGAATGTAAGAAATAAGTAGCGGCACATGTTGGCTCACCCCTGTAATCCAGCACT 670
Db      405  -----ValAlaHisAlaCysAsnProSerThr 413
QY      669  TTGGATCCCAAGTGTAGGAGATTGTTGAAGCCAGGAGTTTAAACAAAGCTGCCCAAC 610
Db      414  LeuGlyGlyGlnAsp-GlyTrpIle-ThrArgSerGlyVal---AspGlnProAspGlnH 432
QY      609  GTAGTAAGACCCAGTCTCTACAAAAATTTTTTAAATACCAAGGCTGTGTGTGACG 550
Db      432  isGlyGluThrProSerLeuThrLysAsnThrLysAsn---ProGlyMetValAlaCysA 451
QY      549  CTCGTAGTCCAGCTACTCTGGGAGGCTG----- 522
Db      451  laCysAsnAlaSerTyrSerGlyGlyArgArgThrAlaThrTrpGluValGluValA 471
QY      521  -----AGTCAGCAGATCCC----- 507
Db      471  laValSerArgAspHisAlaThrAlaLeuGlnProGlyGlnGlnSerGluThrProSerG 491
QY      506  -----TTGAGCTCAGGAGGTTGAGGCTGAGTGAAC 476
Db      491  lnLysLysLysLysLysLysAsnCysLeuAsnProGlyGlyGlyGlyCysSerGluP 511
```

```
QY      475  TATGTTGCACCACTGTCTCATGCTGGTGACAGCAAGACCTCTCTAAACAAC 416
Db      511  roArgLeuHisHisCysThrProValTrpValThrGluArgAspSerValSerLysGlnG 531
QY      415  AACAAAAAAACCCAA 399
Db      531  lnGlnGlnLysThrGln 536
RESULT 4
AAU31513
ID      AAU31513 standard; protein; 215 AA.
AC      AAU31513;
DT      18-DEC-2001 (first entry)
DE      Novel human secreted protein #2004.
KW      Human; vaccination; gene therapy; nutritional supplement;
KW      stem cell proliferation; haematopoiesis; nerve tissue regeneration;
KW      immune suppression; immune stimulation; anti-inflammatory; leukaemia.
OS      Homo sapiens.
FN      WO200179449-A2.
PD      25-OCT-2001.
PF      16-APR-2001; 2001WO-US008656.
PR      18-APR-2000; 2000US-00552929.
PR      26-JAN-2001; 2001US-00770160.
XX      (HYSE-) HYSEQ INC.
PA      Tang YT, Liu C, Drmanac RT;
PI      WPI; 2001-611725/70.
XX      Nucleic acids encoding a range of human polypeptides, useful in genetic
XX      vaccination, testing and therapy.
PS      Claim 20; Page 452; 765pp; English.
```

The invention relates to novel human secreted polypeptides. The polypeptides and antibodies to the polypeptides are useful for determining the presence of or predisposition to a disease associated with altered levels of polypeptide. The polypeptides are also useful for identifying agents (agonists and antagonists) that bind to them. Cells expressing the proteins are useful for identifying a therapeutic agent for use in treatment of a pathology related to aberrant expression or physiological interactions of the polypeptide. Vectors comprising the nucleic acids encoding the polypeptides and cells genetically engineered to express them are also useful for producing the proteins. The proteins are useful in genetic vaccination, testing and therapy, and can be used as nutritional supplements. They may be used to increase stem cell proliferation; to regulate haematopoiesis; and in bone, cartilage, tendon and/or nerve tissue growth or regeneration; immune suppression and/or stimulation; as anti-inflammatory agents; and in treatment of leukaemias. AAU29510-AAU33304 represent the amino acid sequences of novel human secreted proteins of the invention

SQ Sequence 215 AA;

Alignment Scores:		
Pred. No.:	1,54e-41	Length: 215
Score:	438.50	Matches: 116
Percent Similarity:	59.9%	Conservative: 14
Best Local Similarity:	53.5%	Mismatches: 60
Query Match:	12.4%	Indels: 28
DB:	4	Gaps: 3

US-10-607-806-1_COPY_7200_9200 (1-2001) x AAU31513 (1-215)

QY 439 TCTGTACCCAGGCGATGACGACAGTGGTGCAACCAATAGGTCACTGACGCTCAACCTCCT 498
 Db 4 SerValThrGlnAlaGlyValGlnTrpCysAsnLeuGlySerLeuGlnProLeuProPro 23
 QY 499 GAGCTCAAGGATCTGCTGACCTCAAGCTCCCAAGTACCTGGATCGACGCGTGCCACA 558
 Db 24 ArgPheArg***PheSerCysLeuSerLeuLeuSerSerTrpAspTrpArgArgProPro 43
 QY 559 CCACCGCTGGCTAAATTAATAAATTTTGTAGAGACTGGGTCTTACTAGCTGGCCAG 618
 Db 44 ProArgProAlaAsn-PheCysIlePheSerArgAspMetAlaPheThrThrLeuAlaAr 63
 QY 619 GCTTCTCTTAATCCTGCTTCAAGC-AATCCTCTTACTCTTGGCATCCCAAGGCTGG 677
 Db 63 GlnValSerAsnSer***ProGlnCysAspProProThrSerAlaSerGlnSerAlaG1 83
 QY 678 GATTACAGGCTGAGCCACCATGTGGGCTACTTATTCTTTTACATTCCATCTTCCAAAT 737
 Db 83 uileThrGlyValSerHisArgAlaTrpProValLeuSerSerProGlnProPhe----- 101
 QY 738 AGAATGTAAGATCCACAGAACAGGAGTACTGCTTATTCTTCTCTTCTTTTGTAGAC 797
 Db 102 -----PhePhePheAspMe 106
 QY 798 AGAGTCTCACTTCATCACTCAACCTCAACCTCGGTTGAGC-----TCAGTGCAC 842
 Db 106 tGluSerHisAlaIleThrGlnAlaGlyVal-GlnTrpArgHisLeuGlySerLeuGlnP 126
 QY 843 CTCTGCTCCGGGTCAAGYATCTCTGCTAAGCTCCTGAGCTGAGTACGCTGGAATTACA 902
 Db 126 roProProProMetPheLys***SerCysLeuSerLeuLeuSerSerTrpAspTrpA 146
 QY 903 AGCGTCACACCATCTTGGCTAATTTTGTATTTT---AGCAGAGATGGGCTTTTA 959
 Db 146 rgArgProProProArgProAlaAsnPhePhe-ValPheLeu**ArgAspGlyValSer 165
 QY 960 CCATGTTGCCAGGCTGCTCAAACTCCTGACCTCAAGTATCTGCTGCTGCTGCTCAGTCTC 1019
 Db 166 ProCys**ProGlyTrpSerArgSerProAspLeuValIleHis-ProProTrpSerPr 185
 QY 1020 CCMAAGTCTGGAATATAGGCTGAGTCACTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 1062
 Db 185 olySerAlaGlyIleThrGlyLeuSerHisCysAlaGlnPro 199

RESULT 5
 AAE35264
 ID AAE35264 standard; protein; 7036 AA.
 XX
 AC AAE35264;
 XX

28-MAY-2003 (first entry)

Human P450RAI-3 protein #1.

XX Human; retinoic acid; vitamin A; cytochrome P450; actinic keratosis;
 KW P450RAI; cancer; psoriasis; acne; ichthyosis; gene therapy; vaccine;
 KW enzyme.

XX Homo sapiens.

OS WO200295034-A2.

XX 28-NOV-2002.

XX 23-MAY-2002; 2002WO-CA000758.

XX 23-MAY-2001; 2001US-0292531P.

XX (CYTO-) CYTOCHROMA INC.

XX Wisniewski J, Petkovich PM, Ramshaw H;

XX WPI: 2003-112152/10.
 DR N-PSDB; AAD53861.
 XX
 PT New isolated P450RAI-3 nucleic acid molecule and polypeptide, useful for
 PT diagnosing, preventing, or treating disorders with aberrant expression or
 PT activity of the P450RAI-3, such as cancer, actinic keratosis, acne and
 PT psoriasis.
 XX
 PS Example 1; Fig 1; 231pp; English.
 XX
 CC The invention relates to retinoic acid metabolising cytochrome P450,
 CC P450RAI polypeptides and polynucleotides. Methods and compositions of the
 CC invention are useful for diagnosing, preventing, ameliorating and/or
 CC treating disorders associated with the aberrant expression or activity of
 CC the P450RAI such as diseases related to vitamin A and retinoic acid
 CC metabolism, e.g. cancer, actinic keratosis, psoriasis, acne and
 CC ichthyosis. The invention is useful in gene therapy and as vaccine. The
 CC present sequence is human P450RAI protein
 XX
 SQ Sequence 7036 AA;

Alignment Scores:
 Pred. No.: 3.83e-40 Length: 7036
 Score: 432.00 Matches: 184
 Percent Similarity: 40.6% Conservative: 70
 Best Local Similarity: 29.4% Mismatches: 160
 Query Match: 12.3% Indels: 212
 DB: 6 Gaps: 26

US-10-607-806-1_COPY_7200_9200 (1-2001) x AAE35264 (1-7036)

QY 1831 TTTATTTTATTT-----ATTATTTTTCAGATAGTCTCGTGTGCGCCAGGCT 1778
 Db 123 PheValLeuPheCysValValLeuPheLeuGluAlaGlySerCysSerAlaAlaHisPro 142
 QY 1777 GGAGTGCAGTGGTGGATCTCGGCTCACTGCAAACTCTGCTCCAGGTTTCAAGCAATTC 1718
 Db 143 GlyValGlnPhe-----Ile-AlaHisCysSerLeuHisLeuLeuAspSerProPr 160
 QY 1717 TCTTGCTCAGCTCCGAGTGGAGTCTACAGGATGATCATCTACACCCAGCTAAT 1658
 Db 160 o---ThrSerAlaSerGlnValAlaSerThrThrGlyValHisHisSerGlnLeuI 179
 QY 1657 TTTTGTATTTTAGTAGAGATAGGTTTCCACATATTCACCCAGGCTGCTCGAATCTCT 1598
 Db 179 ePheAsnCysMetSerLysGlnGlyLeu-----ThrValLeuProArgLeu 194
 QY 1597 CACCTCAAGTATCCACCACCTTGGACTTCCAAAGTACTGTGGTTACAGGCTTGAGTCA 1538
 Db 194 uValSerAsnSerSerGlnAlaLeuAlaSerGlnSerAlaArgIleThrGlyValSerHi 214
 QY 1537 CTGTGCCCACTGAGTTATTTTAAATTAATATGAACAAACATCTCTTAATGCGAGTTTCT 1478
 Db 214 eHisGlyArgProGluAspHisMetLeuPheAspSerHisLeu----- 228
 QY 1477 TGTCTTAGTGCACCAATTAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAA 1418
 Db 229 -----GlySerAspCys----- 232
 QY 1417 CTCTGGGAGAGGTTTCATAGCTTTTCATCAGATTTTGTAAATTTGGCCTT---CTAAAGTG 1361
 Db 233 -----TrpLeuArgGlyPhePheGln-----GlyLeuAsnLeuArgTr 245
 QY 1360 GAGCAGGAGGATGTATAAGCAAAATTAAG-----CCATATTATTTCAGTGAC 1316
 Db 245 pLeuHisAsnSerValAsnArgValIysThrIleAspLeuCysProLeuAsnGluLeuTh 265
 QY 1315 TAAATAGTAAACAATCAAAAACAGTAAACAATAAAGTTTGGAGCAATATGAAAGTCA 1256
 Db 265 rValTrpTrpTrpTrpTrpTrpTrpTrpTrpTrpTrpTrpTrpTrpTrpTrpTrpTrpTrp 281
 QY 1255 TAACAAGGTTGATAAGTGAATAAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1196


```

Db 71 SerLeuGluLeuLeuAlaSerGlyAspProProSerGlyAspProAspLeuArg***Leu 90
QY 661 GCATCCCAAGTGTGGGATTACAGGGGTGAGCCACCATGTGGCTACTATTCTTTTA 720
Db 91 AlaPheGlnSerLeuGlyLeuThrGlyValThrHisHis-AlaGlyGlnIlePheLeuPh 110
QY 721 CATTCATCTTTCCCAATAGATGTAGATCCACAGAACAGGAGTACTGCCTATTCTT 780
Db 110 eCysLeuGlu----- 113
QY 781 CTTTCTTTTGTGACAGAGTCTCATTCATCACTCAACCTCCG----- 826
Db 114 -----ThrGluSerProSerValAlaGlnValGlyValG1 125
QY 827 -----TTCACTCACTCAACCTCTGCTCCCGGTTCAAGYGATTCTCTGCT 876
Db 125 nTrpCysAsnLeuGlySerLeuGlnProLeuProProGlySerGlu***PheSerCysLe 145
QY 877 AAGCTCTCTGAGTGTGAATTAACAAGCGTGCACACCATCTTGGCTAAATTTTGTGA 936
Db 145 uGlyLeuProSerSerTrpAspTyrArgHisAlaLeuThrHisProAlaAsn-PheSerI 165
QY 937 TTTTACGACAGATGGGTTTACCATGTGTCAGGCTGGTCTCAAACTCTGACCTC 995
Db 165 lePheSerArgAspGlyValSerProSerTrpProGlyTrpSerGlnMetProAspLeu 184

```

RESULT 10

ADK36978
ID ADK36978 standard; protein; 232 AA.

AC ADK36978;

DT 06-MAY-2004 (first entry)

DE Novel human polypeptide SeqID9060.

XX antiarthritic; antiparkinsonian; neuroprotective; nootropic;
KW immunosuppressive; cytostatic; antipsoriatic; antiinflammatory;
KW antibacterial; antiviral; antifungal; antiparasitic; gene therapy;
KW arthritis; Parkinson's; Alzheimer's; autoimmune disease; cancer;
KW psoriasis; inflammatory bowel disease; infection; bacteria; virus;
KW fungus; parasite; human.

XX Homo sapiens.

EH Key Location/Qualifiers

FT Misc-difference 1..232

FT /label= OTHER

FT /note= "OTHER= All xaa's in this sequence are unknown
amino acids or the site of a stop codon within the DNA
sequence"

FT WO200216439-A2.

FN 28-FEB-2002.

PD 05-MAR-2001; 2001WO-US004941.

PF 07-MAR-2000; 2000US-00519705.

PR 19-MAY-2000; 2000US-00574454.

XX (HYSE-) HYSEQ INC.

PI Tang YT, Liu C, Drmanac RT;

XX WPI; 2002-280918/32.

XX Isolated polynucleotide encoding bone marrow derived polypeptides useful
for treating, e.g., Parkinson's, Alzheimer's, cancer, arthritis, Crohn's
disease, and inflammatory bowel disease.

PS Claim 20; SEQ ID NO 9060; 504pp; English.

XX

CC This invention relates to a novel isolated polynucleotide comprising a
CC nucleotide sequence selected from one of 1680 sequences, a mature protein
CC coding portion of them, an active domain of them and their complementary
CC sequences. The invention may be useful for the production of compounds
CC with an antiarthritic, antiparkinsonian, neuroprotective, nootropic,
CC immunosuppressive, cytostatic, antipsoriatic, antiparasitic, antiinflammatory,
CC antibacterial, antiviral, antifungal or antiparasitic activity. In
CC addition, the disclosed sequences may be useful for gene therapy. The
CC polypeptides or their antibodies are useful for treating many diseases
CC such as arthritis, Parkinson's, Alzheimer's, autoimmune diseases, cancer,
CC psoriasis, inflammatory bowel disease and infections caused by bacteria,
CC viruses, fungi or parasites. The present sequence is that of a human
CC polypeptide of the invention.

XX Sequence 232 AA;

Alignment Scores:

Pred. No.: 1.36e-37 Length: 232
Score: 405.00 Matches: 108
Percent Similarity: 56.1% Conservative: 11
Best Local Similarity: 50.9% Mismatches: 52
Query Match: 11.5% Indels: 42
DB: 5 Gaps: 5

US-10-607-806-1_COPY_7200_9200 (1-2001) x ADK36978 (1-232)

```

QY 400 TGGGTTTTTTTGTGTTGTATTAGACAGAGGGTCTTGCTGTGTCACCCAGGCATGACGA 459
Db 52 TrpProPhePhe-----PhePheGlnThrGluSerArgSerValThrGlnAlaGlyVal 69
QY 460 CAGTGTGTCAACCATAGTCACTGCAGCTCAACCTCTGAGCTCAAGGGATCTGCTGAC 519
Db 70 Gln**CysAsnLeuSerSerLeuGlnProLeuProGlyLeuLysGlnPheSerArg 89
QY 520 CTCAGCTCCCAAGTAGCTGGGACTACGAGCGTGCCACACACCCCTGGCTTAATTAATA 579
Db 90 LeuSerLeuProSerSerTrpAspTyrArgHisMetProProArgProAlaSer--Phe 108
QY 580 AATTTTGTGAGAGCTGGGTCTTACTAGTGTGGCCAGGCTTCTTAACTCTGTCGC 638
Db 109 ValPheLeuValGluThr-GlyPheThrMetLeuAlaArgLeuValTrpAsnSer**Pr 128
QY 639 TTCAAGCAATCTCTCTACCTTGGCATCCCAAGGTGCTGGGATTACAGGGGTGAGC---- 693
Db 128 oSerCysAspProProIleSerAlaSerGlnAsnAlaGlyIleThrGly**SerProSe 148
QY 694 -----CACCATGTGGCTACTTATTCTTTTACATTTCCATCTTTCCAATAGAATG 743
Db 148 rLeuAlaProHisHisProArgLeu----- 156
QY 744 TAAGATCCACAGAACAGGAGTACTGCTATTCTTCTTCTTTTGTGAGACAGAGTC 803
Db 157 -----PhePheGluThrLysE 163
QY 804 TCATCTTCATCACTCAACCTCCGTTTCAGC-----TCACTGCAACCTCTGC 848
Db 163 rGlySerValThrGlnAlaGlyVal-GlnTrpHisAsnProThrSerGlnGlnProLeuP 183
QY 849 CTCCTGGGTTCAAGYGATTCTCTGCTTAAGCCCTCTGAGTAGCTGGAATTAACAAGCGTG 908
Db 183 roProGlyPheLysGlnLeuSerArgLeuSerLeuSerSerSerTrpAsp**ArgTyrA 203
QY 909 CACCACCATCTGGCTAAATTTTGTATTTTGTAGCAGAGATGGGGTTTACCATGTGTC 968
Db 203 laProSerCysLeuAlaAsn-TyrCysIlePheSerArgAspArgAlaSerProCysTrp 222
QY 969 CAGGCTGTCTCAAACTCTCTCAACCTCAAG 998
Db 223 SerGlyTrpSer**ThrProAspLeuArg 232

```

RESULT 11

ADK36936

```

ID XX ADK36936 standard; protein; 302 AA.
AC AC
XX AC
XX ADK36936;
DT 06-MAY-2004 (first entry)
XX
DE Novel human polypeptide SeqID9018.
XX
XX antiarthritic; antiparkinsonian; neuroprotective; nootropic;
KW immunosuppressive; cytostatic; antipsoriatic; antiinflammatory;
KW antibacterial; antiviral; antifungal; antiparasitic; gene therapy;
KW arthritis; parkinson's; Alzheimer's; autoimmune disease; cancer;
KW psoriasis; inflammatory bowel disease; infection; bacteria; virus;
KW fungus; parasite; human.
XX
OS Homo sapiens.
XX
XX Key Location/Qualifiers
FH Misc-difference 1..302
FT /label= OTHER
FT /note= "OTHER= All Xaa's in this sequence are unknown
FT amino acids or the site of a stop codon within the DNA
FT sequence"
XX
XX WO200216439-A2.
XX
XX 28-FEB-2002.
XX
XX 05-MAR-2001; 2001WO-US004941.
XX
XX 07-MAR-2000; 2000US-00519705.
XX
XX 19-MAY-2000; 2000US-00574454.
XX
XX (HYSE-) HYSEQ INC.
XX
XX Tang YT, Liu C, Drmanac RT;
XX
XX WPI; 2002-280918/32.
XX
XX Isolated polynucleotide encoding bone marrow derived polypeptides useful
PT for treating, e.g., Parkinson's, Alzheimer's, cancer, arthritis, Crohn's
PT disease, and inflammatory bowel disease.
XX
XX Claim 20; SEQ ID NO 9018; 504pp; English.
XX
XX This invention relates to a novel isolated polynucleotide comprising a
CC nucleotide sequence selected from one of 1680 sequences, a mature protein
CC coding portion of them, an active domain of them and their complementary
CC sequences. The invention may be useful for the production of compounds
CC with an antiarthritic, antiparkinsonian, neuroprotective, nootropic,
CC immunosuppressive, cytostatic, antipsoriatic, antiinflammatory,
CC antibacterial, antiviral, antifungal or antiparasitic activity. In
CC addition, the disclosed sequences may be useful for gene therapy. The
CC polypeptides or their antibodies are useful for treating many diseases
CC such as arthritis, Parkinson's, Alzheimer's, autoimmune diseases, cancer,
CC psoriasis, inflammatory bowel disease and infections caused by bacteria,
CC viruses, fungi or parasites. The present sequence is that of a human
CC polypeptide of the invention.
XX
XX Sequence 302 AA;
SQ
Alignment Scores:
Pred. No.: 5.84e-37 Length: 302
Score: 400.00 Matches: 110
Percent Similarity: 53.1% Conservative: 20
Best Local Similarity: 44.9% Mismatches: 72
Query Match: 11.4% Indels: 43
DB: 5 Gaps: 5
US-10-607-806-1_COPY_7200_9200 (1-2001) x ADK36936 (1-302)
QY 397 GTTTGGGTTTTTTTTTTTGTGTGT-----TTAGAGACAGGGTCTTGCTCT 441
||||| ||||||| ||| |||||

```

```

Db 81 ValCysThrPhePheIleTyrCysProPheLeuPhePheLeuArgThrGlyPheSerAla 100
Qy 442 GTACCCAGGCGATGAGCAGACAGTGGTCAACATAGGTCACTGCAGCCTCAACCTCCTGAG 501
|||:||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 101 LeuSerGlnAlaValValGlnTrpArgAspLeuSerSerLeuGlnProProProGly 120
|||:||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Qy 502 CTCAGGGGATCTGCTGACCTCAGCTCCCAAGTAGCTGGGACTACGAGCGGTGCACCA 561
|||:||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 121 LeuLysArgPheSerCysLeuSerLeuProSerSerTrpAspHisArgHisValProPro 140
|||:||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Qy 562 CGCTGGCTAATTAATAAATTTTGTAGAGACTGGGTCTTACTAGTTGGCCAGGCT 621
|||:||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 141 HisProAlaAsn-PheCysIleLeuMet***ArgTrpGlyPheThrMetPheAlaArgMe 160
|||:||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Qy 622 TGTCTTAAACTCTGGCTTCAAGCAATCTCTACCTTGGCATCCCAAGTGTGGGATT 681
|||:||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 160 tThrCysAsp-----LeuLeu--ThrSerAlaSerGlnThrValArgIle 174
|||:||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Qy 682 ACAGGGGTGAGCCACCATGTGCGGTACTTATTCTTTTACATTCATCTTTCCCAATAGAA 741
|||:||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 175 ThrGlyValSerHisHisThrArgLeu----- 183
Qy 742 TGTAGATCCACAGACAGGAGTACTGCCCTATTTT-----CTTCCTTTC 786
|||:||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 184 -----IleTyr***TyrPhe***Ala***SerAspIlePhePhe 196
|||:||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Qy 787 TTTTGTGAGCAGAGTCTCATTCACTCACTCACTCCGTTTCAGC----- 832
|||:||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 197 PhePheGluLysGluSerCysCysValGlyGlnAlaGlyVal-GlnTrpHisAspLeuGln 216
|||:||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Qy 833 -TCACTGCAACCTCTGCTCCCGGTTCAGAGGATTTCTCTGCTTAAGCCTCCTGAGTAG 891
|||:||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 216 ySerLeuGlnProProProSerGlyLeuLysGlnPheSerCysLeuAsnLeuProSerSe 236
|||:||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Qy 892 CTGGAATTACAAGCGTGCACCATGCTGGCTAAATTTTGTATTTTATTTTACGAGATG 951
|||:||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 236 rTrpAspHisLysArgThrProHisProGlySer-PheCysIlePheSerArgAspA 256
|||:||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Qy 952 GGGTTTACCATGTTGCCAGGCTGGTCTCAAACTCCTGAGCTCAAGTATCTGCTGCC 1011
|||:||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 256 rgValSerProTrpGlyProGlyTrpPhe***ThrProAspLeuLysArgSerAlaLeu 276
|||:||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Qy 1012 TCAGTCTCCCAAAGTGTGGAAATTATAGGCGTGGTCACTGCTGCTGCCGCGGATTACTGTC 1071
|||:||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 276 euGlyLeuProLysCysCysAspTyrArgArgGluProProArgProAlaGluValThrI 296
|||:||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Qy 1072 TATTT 1076
|||:|||||
Db 296 lePhe 297
|||:|||||
XX
XX RESULT 12
XX ID ABG68738 standard; protein; 361 AA.
XX AC ABG68738;
XX
XX 07-OCT-2002 (first entry)
XX
XX Human prostate specific protein DEX0293_107.
XX
XX Human; prostate specific nucleic acid; PSA; prostate cancer; PSP;
KW prostate specific protein; cytostatic; non-cancerous prostate disease;
KW gene therapy; cancer; immunostimulant; vaccine.
XX
XX Homo sapiens.
XX
XX WO200255735-A2.
XX
XX 18-JUL-2002.
XX
XX 27-NOV-2001; 2001WO-US044363.
XX
XX 27-NOV-2000; 2000US-0253176P.
XX
XX

```


XX (DIAD-) DIADEXUS INC.
 XX Salceda S, Macina RA, Recipon H, Cafferkey R, Ali S, Sun Y;
 XX Liu C, Chen S;
 XX WPI; 2002-557831/59.
 DR N-PSDB; ABK97631.
 XX New prostate specific genes, useful for treating or diagnosing cancer, or
 PT useful as vaccines for treating cancer, particularly prostate cancer, in
 PT a patient.
 XX Claim 11; Page 207-208; 212pp; English.
 XX The invention relates to a new isolated prostate-specific nucleic acid
 CC (PSNA) molecule comprising the cDNA sequences appearing as ABK97574-
 CC ABK97642 which encode prostate specific proteins appearing as ABG68701-
 CC ABG68746, or a sequence hybridising to a PSNA or which has 60% sequence
 CC homology with a PSNA. Also included are a method of determining the
 CC presence of a PSNA in a sample, a vector comprising the PSNA, a host cell
 CC comprising the vector, producing the polypeptide encoded by the PSNA, a
 CC method of determining the presence of a PSP in a sample, diagnosing and
 CC monitoring the presence and metastases of prostate cancer in a patient, a
 CC kit for detecting a risk of cancer or presence of cancer in a patient
 CC (the kit comprising a means for determining the presence of the PSNA or
 CC PSP in a sample of a patient) and a vaccine comprising the polypeptide or
 CC the nucleic acid encoding the polypeptide. The PSNA, PSP and anti-PSP
 CC antibody are useful for diagnosing and treating cancer in a patient (e.g.
 CC by gene therapy). The nucleic acid molecule and polypeptide are also
 CC useful as vaccines for treating cancer, particularly prostate cancer and
 CC non-cancerous prostate diseases. The present sequence is a PSP of the
 CC invention
 XX Sequence 361 AA;
 SQ
 Alignment Scores:
 Pred. No.: 6.3e-37 Length: 361
 Score: 400.00 Matches: 105
 Percent Similarity: 55.8% Conservative: 11
 Best Local Similarity: 50.5% Mismatches: 55
 Query Match: 11.4% Indels: 38
 DB: 5 Gaps: 5
 US-10-607-806-1_COPY_7200_9200 (1-2001) x ABG68738 (1-361)
 QY 439 TCTGTACCCAGGATGACGACAGTGGTGCACACATAGGTCTACCTGACCTCACTCT 498
 Db 172 SerValAlaGlnAlaGlyValGlnTrpCysAsnLeuGlySerLeuGlnProLeuPro 191
 QY 499 GAGCTCAAGGGATCTGTGACCTCAGCTCCCAAGTAGCTGGGACTACGAGCGGCACA 558
 Db 192 ArgPheArgArgPheSerCysLeuSerLeuSerTrpAspHisArgHisAlaPro 211
 QY 559 CCACGCCCTGGCTAAATAAANA-----AATTTTGTAGAGCTGGGTCTTACTACGTT 612
 Db 212 ProCysLeuAlaAsnPheLeuPhePheLysPheLeuValAspGln-SerPheThrMetLe 231
 QY 613 GGCAGGCTGTCTTAACCTCTG-GCTTCAAGCAATCTCTTACCTTGGCATCCCAAG 671
 Db 231 uAlaArgLeuValLeuAsnSerAlaProSerGlyAspLeuProAlaProAlaSerGlnSe 251
 QY 672 TGCTGGGATTACAGGGTGAGCCAC----- 696
 Db 251 rAlaGlyIleThrGlyValArgHisCysThrTrpProLysSerPheLysPheAlaAspSe 271
 QY 697 -CATGTGCGGTACTTATTCTTTTACATTCCTTTTCCATAGATGATGATCCACAG 755
 Db 271 rHisIleGlyLeu-----AlaPheHisPheAlaPhe----- 281
 QY 756 AACAGGGATTACTGCTATTTCTTCTTCTTTTGGAGACAGAGTCTCATCTCATCAC 815
 Db 282 -----PhePhePhePhePhePhePheAlaValAlaSerHisProIleAl 296

QY 816 CTCAACCTCCGTCCTGCTCTCCCGGTTCA 860
 Db 296 aGlnAlaGlyVal-GlnTrpArgAspLeuGlySerLeuGlnProProProGlyPheL 316
 QY 861 AGYGATTCCTCTGCTAAGCTCTGAGTAGCTGGAATTACAAGCGTGACACACCAAGCT 920
 Db 316 yGlnPheLeuCysLeuSerLeuProGlySerTrpAspTyArgAlaProProArg 336
 QY 921 TGGCTTAATTTTCTTATTTTACAGAGATGGGCTTTTACCATGTTGCCAGGCTGGTCT 980
 Db 336 lnAlaAsn-PheCysIlePheSerArgAspGlyValSerProCysTrpThrGlyTrpSer 355
 QY 981 CAAACTCTCTGACCTCAAG 998
 Db 356 GlnThrProAspLeuArg 361

RESULT 13

ABJ01057
 ID ABJ01057 standard; protein; 449 AA.
 AC ABJ01057;
 XX 28-NOV-2002 (first entry)
 XX Ovary cell-specific amino acid sequence 3.
 DE Ovary cell; neoplastic ovary cell; ovary specific nucleic acid;
 KW ovary specific protein; ovarian cancer; breast cancer; vaccine;
 KW gene therapy.
 XX Homo sapiens.
 XX WO200238606-A2.
 XX 16-MAY-2002.
 XX 07-NOV-2001; 2001WO-US046459.
 PR 08-NOV-2000; 2000US-0246640P.
 XX (DIAD-) DIADEXUS INC.
 PI Sun Y, Recipon H, Salceda S, Liu C;
 XX WPI; 2002-519297/55.

Polypeptide and polynucleotides present in normal and neoplastic ovary cells, useful for identifying, monitoring, staging, diagnosing, preventing and treating ovarian cancer, and non-cancerous disease states in the ovary.

Claim 11; Page 207-209; 247pp; English.

The invention comprises amino acid and DNA sequences which are present in normal and neoplastic ovary cells. The DNA and protein sequences of the invention are useful for determining the presence of an ovary specific nucleic acid or an ovary specific protein in a sample. The DNA and protein sequences of the invention are useful for diagnosing and monitoring the presence and metastasis of ovarian cancer and breast cancer. Amino acids ABJ01055 - ABJ01155 represent the ovary cell specific protein sequences of the invention

Sequence 449 AA;

Alignment Scores:

Pred. No.: 6.9e-37 Length: 449
 Score: 400.00 Matches: 103
 Percent Similarity: 58.1% Conservative: 26
 Best Local Similarity: 46.4% Mismatches: 62
 Query Match: 11.4% Indels: 31
 DB: 5 Gaps: 7

Db 311 -GlnAlaGlyValGlnTrpProAsnLeuGlySerLeuGlnProLeuProProGlyLeuLy 330
QY 862 GYGATTCTCCTGCCTAAGCCTCCTAGTAGCTGGAATTACAAAGCGTGCCACCACCATGCTT 921
Db 330 sArgPheSerCysLeuSerLeuProSerSerTrpAspTyrGlyHisLeuProProHisPr 350
QY 922 GGCTAAATTTTGTATTTTAGCAGAGATGGGGTTTACCATGTTGCCACGCTGCTTC 981
Db 350 oAlaAsn-PheCysIlePheIleArgGlyGlyValSerProTyrLeuSerGlyTrpSerG 370
QY 982 AAATCCTGACCTCAAG 998
Db 370 InThrProAspLeuArg 375

Search completed: February 8, 2006, 10:00:48
Job time : 387.279 secs

GenCore version 5.1.7
Copyright (c) 1993 - 2006 Bioceleration Ltd.

OM nucleic - protein search, using frame_plus_n2p model

Run on: February 8, 2006, 10:22:47; Search time 19.8358 Seconds
(without alignments)
3105.549 Million cell updates/sec

Title: US-10-607-806-1-G7328_COPY_7000_7700

Perfect score: 1253

Sequence: 1 gtcgtgctactgctgtccag.....tcgagctcaactcctcgag 701

Scoring table:

BLOSUM62
Xgapop 10.0, Xgapext 0.5
Ygapop 10.0, Ygapext 0.5
Fgapop 6.0, Fgapext 7.0
Delop 6.0, Delext 7.0

Searched: 2443163 seqs, 439378781 residues

Total number of hits satisfying chosen parameters: 4886326

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Command line parameters:

-MODEL=frame+...n2p.model -DEV=xlp
-Q=/abss/ABSSWEB_epool/MARTINI067806/runat_08022006_092016_27646/app_query.fasta_1
-DB=A Geneseq -QFMT=fastan -SUFFIX=n2p.rag -MINMATCH=0.1 -LOOPCL=0 -LOOPEXT=0
-UNITS=bits -START=1 -END=-1 -MATRIX=blosum62 -TRANS=human40.cdi -LIST=45
-DOALIGN=200 -THR SCORE=pct -THR MAX=100 -THR MIN=0 -ALIGN=15 -MODE=LOCAL
-OUTFMT=ptc -NORM=ext -HEAPSIZE=500 -MINLEN=0 -MAXLEN=200000000 -HOST=abs02p
-USER=MARTINI067806 @CGN_1_1_605 @runat_08022006_092016_27646 -NCPU=6 -ICPU=3
-NO_WMAP -NEG_SCORES=0 -WAIT -DSPBLOCK=100 -LONGLOG -DEV_TIMEOUT=120
-WARN_TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAEXT=0.5 -FGAPOP=6 -FGAPEXT=7
-YGAPOP=10 -YGAEXT=0.5 -DELOP=6 -DELEXT=7

Database :

A_Geneseq_21.*
1: geneseqp1980s.*
2: geneseqp1990s.*
3: geneseqp2000s.*
4: geneseqp2001s.*
5: geneseqp2002s.*
6: geneseqp2003as.*
7: geneseqp2003bs.*
8: geneseqp2004s.*
9: geneseqp2005s.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	260	20.8	148	5 AAU78668	Aau78668 Human Pho
2	258	20.6	148	2 AAR26410	Aar26410 Human pho
3	258	20.6	148	2 AAR63054	Aar63054 Human PLA
4	258	20.6	148	5 AAU78667	Aau78667 Human Pho
5	258	20.6	148	6 ABR44232	Abr44232 Human sec
6	258	20.6	148	6 ABU63122	Abu63122 Human gro
7	258	20.6	148	8 ADI35083	Adi35083 Human PLA
8	258	20.6	148	8 ADJ09984	Adj09984 Human pho
9	258	20.6	156	3 AAB54059	Aab54059 Human pan

10	258	20.6	194	8 ABM83696	Abm83696 Human dia
11	258	20.6	217	4 ABG11912	Abg11912 Novel hum
12	244	19.5	145	8 ADO28599	Ado28599 Human PA2
13	229	18.3	79	3 AAB54421	Aab54421 Human pan
14	222	17.7	146	5 ABB08153	Abb08153 Murine PL
15	216	17.2	124	2 AAR63059	Aar63059 Rat PLA2
16	216	17.2	146	2 AAR63052	Aar63052 HPLA2-8.
17	209	16.7	145	3 AAR04821	Aar04821 D-helix d
18	209	16.7	145	3 AAY88294	Aay88294 Bovine PL
19	204	16.3	146	3 ADY72838	Ady72838 Pig phosp
20	204	16.3	146	3 ADY72839	Ady72839 Pig phosp
21	204	16.3	146	3 ADY72840	Ady72840 Pig phosp
22	200	16.0	145	1 AAP94659	Aap94659 Amino aci
23	195	15.6	146	9 ADY72845	Ady72845 Pig phosp
24	195	15.6	146	9 ADY72844	Ady72844 Pig phosp
25	195	15.6	146	9 ADY72841	Ady72841 Pig phosp
26	195	15.6	146	9 ADY72843	Ady72843 Pig phosp
27	195	15.6	146	9 ADY72842	Ady72842 Pig phosp
28	143	11.4	133	6 ADB19008	Adb19008 Gamma-tai
29	130.5	10.4	110	5 ABB06221	Abb06221 Phosphoen
30	128	10.2	25	9 ADY38233	Ady38233 Human CPP
31	125.5	10.0	74	4 AAO11433	Aao11433 Human pol
32	125	10.0	7052	6 AAE35266	Aae35266 Human P45
33	124.5	9.9	80	5 AAM47619	Aam47619 Human chl
34	123	9.8	100	4 AAO04010	Aao04010 Human pol
35	122	9.7	118	1 AAP90076	Aap90076 Phospholi
36	118	9.4	150	4 AAO01959	Aao01959 Human pol
37	117	9.3	53	4 AAO11877	Aao11877 Human pol
38	117	9.3	53	5 ABF03853	Abf03853 Human ORF
39	116	9.3	115	4 AAM83664	Aam83664 Human imm
40	114.5	9.1	125	4 AAU30406	Aau30406 Novel hum
41	114.5	9.1	127	4 AAU21236	Aau21236 Human nov
42	114	9.1	61	4 AAO12650	Aao12650 Human pol
43	114	9.1	118	3 AAY67383	Aay67383 Acanthin
44	112.5	9.0	475	8 ADX94147	Adx94147 Plant ful
45	112	8.9	122	4 AAO12468	Aao12468 Human pol

ALIGNMENTS

RESULT 1

AAU78668

ID AAU78668 standard; protein; 148 AA.

AC AAU78668;

XX

DT 18-JUN-2002 (first entry)

XX

DE Human Phospholipase A2, groupIB (PLA2G1B) variant.

XX

KW Human; SNP; single nucleotide polymorphism; pancreatitis; enzyme;

KW pancreatic cancer; Phospholipase A2 groupIB; PLA2G1B; gene therapy;

KW haplotype; genotype; chromosome 12q23-q24.1; transgenic; drug screening.

XX Homo sapiens.

OS

XX

Key Location/Qualifiers

FT Misc-difference 122

FT /note= "Wild-type Arg substituted by His"

XX WO200212562-A2.

XX

PD 14-FEB-2002.

XX

PF 06-AUG-2001; 2001WO-US024663.

XX

PR 04-AUG-2000; 2000US-0223179P.

XX

PA (GENA-) GENAISSANCE PHARM INC.

XX

PI Kazemi A, Klieb SE, Koshy B;

XX WPI; 2002-303982/34.

DR

Novel isolated human Phospholipase A2, Group IB pancreas polynucleotide, for therapeutic purposes, for studying expression and function of the polynucleotide and for expressing the phospholipase protein.

Claim 30; Page; 51pp; English.

The invention relates to an isolated human Phospholipase A2, Group IB (pancreas) (PLA2G1B) polynucleotide comprising a sequence which is a polymorphic variant for a reference sequence for the PLA2G1B gene or its fragment, or a polymorphic variant of a reference sequence for a PLA2G1B cDNA or its fragment. Also included are haplotyping/genotyping the PLA2G1B gene of an individual, predicting the haplotype pair for the PLA2G1B gene of an individual, identifying an association between a trait and at least one haplotype or haplotype pair of the PLA2G1B gene, an isolated genotyping oligonucleotide for detecting a polymorphism in the PLA2G1B gene, a recombinant non-human organism transformed or transfected with the PLA2G1B sequence, where the organism expresses a PLA2G1B protein encoded by the first nucleotide sequence or by the polymorphic variant sequence, an isolated polypeptide comprising a sequence which is a polymorphic variant of a reference sequence for the PLA2G1B protein or its fragment, an anti-PLA2G1B monoclonal antibody, screening for drugs targeting PLA2G1B, a computer system for storing and analysing polymorphism data for the PLA2G1B gene and a genome anthology for PLA2G1B gene. The PLA2G1B variant is useful in studying the expression and function of PLA2G1B, and in expressing PLA2G1B protein for use in screening for candidate drugs to treat diseases related to PLA2G1B activity (e.g. pancreatitis and pancreatic cancer) and for therapeutic purposes. The transgenic organism is useful for studying expression of the PLA2G1B isogenes in vivo, for in vivo screening and testing of drugs targeted against PLA2G1B protein, and for testing the efficacy of therapeutic agents and compounds in a biological system. The antibody is useful for studying the effect of the variation on the biological activity of PLA2G1B as well as on the binding affinity of candidate drugs targeting PLA2G1B. The PLA2G1B gene is located on chromosome 12q23-q24.1. The present sequence is the PLA2G1B protein variant. Note: The present sequence is not shown in the specification but was created by the indexer using the PLA2G1B sequence appearing as AAU78667 and the information in figure 3

Sequence 148 AA;

Alignment Scores:

Pred. No.:	5.63e-23	Length:	148
Score:	260.00	Matches:	60
Percent Similarity:	54.5%	Conservative:	13
Best Local Similarity:	44.8%	Mismatches:	26
Query Match:	20.8%	Indels:	35
DB:	5	Gaps:	4

US-10-607-806-1-G7328_COPY_7000_7700 (1-701) x AAU78668 (1-148)

QY 113 ATTTCTCTGAGCATCTTGGCGA-----AGT 139

Db 19 lIeSerProArgAlaValTrpGlnPheArgLysMetIleLysCysValIleProGlySer 38

QY 140 CATCCCCACCTTG-----TTCTCTGAGAGTGGCGAGCTGAGGCGCTG 178

Db 39 AspProPheLeuGluTyrAsnAsnTyrGlyCysTyrCysGlyLeuGly-GlySerGlyth 58

QY 179 ACCTATTGCTCTGCACCTTACTCTCTATCTCAGCTCTCCCTCCCACTTTCCAGGTGCTGCCA 238

Db 58 rProValAsp-----GluleuAspLysCysCysG1 68

QY 239 GACACCAACCTGCTAYGACACGAGCCAGAGCTGAGCAGCTGTAAATTTCTGCTGGA 298

Db 68 nThiGAspAnCysTyrAspGlnAlaLysLysLeuAspSerCysLysPheLeuLeuAs 88

QY 299 CANNCCGTACACCCACCTATTCTGCTCTGCTGCTCGGCAATACCTGTGTAGCAG 358

Db 88 pAsnProTyrThrIsthrTyrSerTyrSerCysSerGlySerAlaIleThrCysSerSe 108

QY 359 TAGG-----TTTATCCCTTCCTTGACCTATGAATCTAGTTGGTTCT 400

Db 108 rLysAsnLysGluCysGluAlaPheIleCysAsnCysAspHisAsnAlaIleCysPh 128

QY 401 CAGTAGCCCGGGGAAATAATAGTAACACAGCCATGAT 440

Db 128 eSerLysAlaProTyrAsnLysAlaHisLysAsnLeuAsp 141

RESULT 2

AA26410

ID AAR26410 standard; protein; 148 AA.

XX AAR26410;

XX 04-MAR-1993 (first entry)

DT Human phospholipase A2 type I.

DE

XX PLA2; rheumatism; osteoporosis; inflammation; immunodeficiency;

XX immunostimulant; vascular smooth muscle cell; synovocyte; chondrocyte;

XX lymphocyte.

XX Homo sapiens.

XX JP04253921-A.

XX 09-SEP-1992.

XX 06-FEB-1991; 91JP-00036789.

XX 06-FEB-1991; 91JP-00036789.

XX (SHIO) SHIONOGI & CO LTD.

XX WPI; 1992-352746/43.

XX Cell growth agents contg. group I phospholipase A2 - used for treating

XX gastric mucous membrane injury, rheumatism, osteoporosis, inflammation,

XX immunodeficiency or vascular injury.

XX Claim 7; Page 2, 9-10; 10pp; Japanese.

XX Group I type phospholipase A2 can specifically bind ca. 90% to cells such

XX as vascular smooth muscle cell, synovocyte, chondrocyte, gastric mucous

XX membrane cell, and spleen-originated lymphocyte. It is used in remedies

XX for gastric mucous membrane injury, rheumatism, osteoporosis,

XX inflammation, immunodeficiency, or vascular injury. Since the agent can

XX increase the growth of spleen-originated lymphocyte, it is utilizable as

XX immunostimulant. Pref. 1 nM or higher concns. of the phospholipase is

XX contacted with cells. Daily dose will pref. be 1mg-100mg given orally or

XX by injection

XX Sequence 148 AA;

Alignment Scores:

Pred. No.:	1e-22	Length:	148
Score:	258.00	Matches:	55
Percent Similarity:	59.8%	Conservative:	6
Best Local Similarity:	53.9%	Mismatches:	12
Query Match:	20.6%	Indels:	29
DB:	2	Gaps:	3

US-10-607-806-1-G7328_COPY_7000_7700 (1-701) x AAR26410 (1-148)

QY 113 ATTTCTCTGAGCATCTTGGCGA-----AGT 139

Db 19 lIeSerProArgAlaValTrpGlnPheArgLysMetIleLysCysValIleProGlySer 38

QY 140 CATCCCCACCTG-----TTCTCTGAGAGTGGCGAGCTGAGGCGCTG 178

Db 39 AspProPheLeuGluTyrAsnAsnTyrGlyCysTyrCysGlyLeuGly-GlySerGlyth 58

QY 179 ACCTATTGCTCTGCACCTTACTCTCTATCTCAGCTCTCCCTCCCACTTTCCAGGTGCTGCCA 238

XX DT 25-SEP-2003 (first entry)
XX DE Human group IIB secreted phospholipase A2.
XX KW Human; group IIB secreted phospholipase A2; virucide; antibacterial;
XX KW cytosolic; antiinflammatory; vasotropic; cerebroprotective; sPLA2;
XX KW phosphatidylglycerol hydrolysis; phosphatidylcholine hydrolysis;
XX KW viral infection; bacterial infection; cancer; inflammatory disease;
XX KW cardiac ischaemia; brain ischaemia; acute lung injury;
XX KW acute respiratory distress syndrome; Crohn's disease; enzyme.
XX OS Homo sapiens.
XX XX US2003073087-A1.
XX PN 17-APR-2003.
XX PD 11-OCT-2001; 2001US-00975456.
XX PF 11-OCT-2000; 2000US-0239491P.
XX PR (LAZD/) LAZDUNSKI M.
XX PA (LAMB/) LAMBEAU G.
XX PA (VALE/) VALENTIN E.
XX XX Lazdunski M, Lambeau G, Valentin E;
XX PI WPI; 2003-567302/53.
XX DR New mammalian secreted group IIF phospholipase A2 or nucleic acid, useful
XX PT for treating or preventing viral or bacterial infections, or cancers, or
XX PT screening inhibitors of the enzyme for treating e.g. inflammatory
XX PT diseases or ischemia.
XX PS Disclosure; Fig 1; 16pp; English.
XX XX The invention describes a mammalian secreted group IIF phospholipase A2
XX CC (sPLA2), which is Ca2+-dependent, maximally active at pH of about 7-8,
XX CC and hydrolyses phosphatidylglycerol versus phosphatidylcholine with about
XX CC a 15-fold preference. The mammalian secreted group IIF sPLA2 protein or
XX CC nucleic acid, or a pharmaceutical composition is useful for treating
XX CC and/or preventing viral infections, bacterial infections, or cancers. The
XX CC inhibitors of sPLA2 or a composition comprising sPLA2 inhibitors is
XX CC useful for treating disease states or disorders involving group IIF
XX CC sPLA2, e.g. inflammatory diseases, cancer, cardiac and brain ischaemia,
XX CC acute lung injury, acute respiratory distress syndrome, or Crohn's
XX CC disease. The enzyme is also useful for screening various chemical
XX CC compounds for treating these diseases. This is the amino acid sequence of
XX CC human group IIB phospholipase A2 used to determine a consensus sequence
XX CC for human sPLA2s
XX SQ Sequence 148 AA;
Alignment Scores:
Pred. No.: 1e-22 Length: 148
Score: 258.00 Matches: 55
Percent Similarity: 59.8% Conservative: 6
Best Local Similarity: 53.9% Mismatches: 12
Query Match: 20.6% Indels: 29
DB: 6 Gaps: 3
US-10-607-806-1-G7328_COPY_7000_7700 (1-701) X ABU63122 (1-148)
QY 113 ATTTCCTGGAACATCTTGGCGA-----AGT 139
Db 19 IleserProArgAlaValTrpGlnPheArgLysMetIleLysCysValIleProGlySer 38
QY 140 CATCCCACTG-----TTCTGAGAGTGGGCGAGGTGAGGGCTG 178
Db 39 AspProPheLeuGluTrpAsnAsnTyrGlyCysTyrCysGlyLeuGly-GlySerGlyTh 58
QY 179 ACCTATTGCTGCACTTACTCTATCTAGCTGTGCCCTCCACCTTTCAGGTGTGCCA 238

Db 58 rProValAsp-----GlLeuAspLysCysGSGI 68
QY 239 GACACATGACAACTGTCTAYGACAGCGCAAGAGCTGACAGCTGTAAATTCGTCTGGA 298
Db 68 nThiHisAspAsnCysTyrAspGlnAlaLysLysLeuAspSerCysLysPheLeuLeuAs 88
QY 299 CMMCCGTACACCCACCACTTATCATCTCTGCTCTGCTCGGCAATCACCTGTAGCAG 358
Db 88 pAsnProTyrThrHisThrTyrSerTyrSerCysSerGlySerAlaIleThrCysSerSe 108
QY 359 TAGG 362
Db 108 rLys 109
RESULT 7
ADI35083
ID ADI35083 standard; protein; 148 AA.
XX AC ADI35083;
XX DT 22-APR-2004 (first entry)
XX DE Human PLA2G1B protein sequence.
XX KW PLA2G1B; fat deposition; leanness; polymorphism;
XX KW non-insulin dependent diabetes mellitus; NIDDM; hyperinsulinemia;
XX KW hypertension; glucose intolerance; dyslipidemia; hypercoagulability;
XX KW microalbuminuria; human.
XX OS Homo sapiens.
XX PN WO2004002295-A2.
XX PD 08-JAN-2004.
XX PF 27-JUN-2003; 2003WO-US020830.
XX PR 27-JUN-2002; 2002US-0392361P.
XX XX (SEQU-) SEQUENOM INC.
XX PI Adam GIR, Langdown ML;
XX DR WPI; 2004-082843/08.
XX DR N-PSDB; ADI35082.
XX PT Diagnosing a predisposition to fat deposition or leanness, useful for
XX PT diagnosing a predisposition to e.g. diabetes or hypertension, comprises
XX PT detecting the presence of a polymorphism in the PLA2G1B nucleic acid from
XX PT the subject.
XX PS Claim 1; SEQ ID NO 2; 91pp; English.
XX XX The invention relates to diagnosing a predisposition to fat deposition or
XX CC leanness in a subject comprising detecting the presence or absence of a
XX CC polymorphic variation associated with fat deposition at a polymorphic
XX CC site in a PLA2G1B nucleotide sequence in a nucleic acid sample from a
XX CC subject, where the presence of the polymorphic variation indicates a
XX CC predisposition to fat deposition in the subject. The polymorphic
XX CC variation is a guanine at position 7328 or thymine at position 9182 of
XX CC the present sequence. The method is useful for diagnosing a
XX CC predisposition to fat deposition or leanness in a subject, and
XX CC consequently for diagnosing a predisposition to non-insulin dependent
XX CC diabetes mellitus (NIDDM) in a subject and conditions such as
XX CC hyperinsulinemia, hypertension, glucose intolerance, dyslipidemia,
XX CC hypercoagulability, or microalbuminuria, which can lead to early
XX CC prescription of preventive measures. The present sequence represents a
XX CC human PLA2G1B protein sequence.
XX SQ Sequence 148 AA;
Alignment Scores:

Pred. No.: 1e-22 Length: 148
Score: 258.00 Matches: 55
Percent Similarity: 59.8% Conservative: 6
Best Local Similarity: 53.9% Mismatches: 12
Query Match: 20.6% Indels: 29
DB: Gaps: 3

US-10-607-806-1-G7328_COPY_7000_7700 (1-701) x ADJ35083 (1-148)

QY 113 ATTTCTCTGAGCATCTTGGCGA-----AGT 139
DB 19 IleserProargAlaValTrpGlnPheArgLysMetIleLysCysValIleProGlySer 38
QY 140 CATCCCCACCTG-----TTCTCTGAGAGTGGCAGGTAGGGCTG 178
DB 39 AspProPheLeuGluTyrAsnAsnTyrGlyCysTyrCysGlyLeuGly-GlySerGly 58
QY 179 ACCTATTGCTCTGCATTTACTCTCTATCTCAGCTGTCCCTCCACTTTCCAGGTGCTGCCA 238
DB 58 rProValAsp-----GluleuAspLysCysCysGly 68
QY 239 GACACATGACAACTGCTATGACAGCCGCAAGAGCTGGACAGCTGTAAATTTCTGTGGA 298
DB 68 nThrHisAspAsnCysTyrAspGlnAlaLysLysLeuAspSerCysLysPheLeuLeuAs 88
QY 299 CAMMCCGTACACCCACACCTATTATCTATCTGCTGTGGCTGGCAATCACCTGTAGCAG 358
DB 88 pAsnProTyrThrHisThrTyrSerTyrSerCysSerGlySerAlaIleThrCysSerSe 108

QY 359 TAGG 362
DB 108 rLys 109

RESULT 8
ADJ09984
ID ADJ09984 standard; protein; 148 AA.
AC ADJ09984;
XX
XX 17-JUN-2004 (first entry)
XX Human phopholipase A2 (PLA2G1B) protein SeqID 2.
XX human; fat reduction; fat deposition; phospholipase A2; PLA2G1B;
KW appetite suppressant; lipase inhibitor; exercise regimen; obesity;
KW non-insulin dependent diabetes mellitus; NIDDM; cardiovascular disorder;
KW hypertension; antidiabetic.
XX Homo sapiens.
XX WO200402296-A2.
XX 08-JAN-2004.
XX 27-JUN-2003; 2003WO-US020831.
XX 27-JUN-2002; 2002US-0392362P.
XX (SEQU-) SEQUENOM INC.
XX Adam GIR, Langdown ML, Denissenko MF, Dennis E, Cantor C;
PI Rubin B;
XX WPI; 2004-071944/07.
DR N-PSDB; ADJ09983.
XX Identifying a candidate therapeutic for fat reduction, useful for
PT treating diabetes, by introducing a test molecule to a system comprising
PT PLA2G1B protein or nucleic acid, and determining the presence of
PT interaction between the compounds.
XX Claim 1; SEQ ID NO 2; 116pp; English.

This invention relates to a novel candidate therapeutic agent useful for fat reduction and disorders related to fat depositions. Specifically, it refers to polymorphic variations in the phospholipase A2 (PLA2G1B) DNA, which is located on chromosome 12q24 and has been associated with central fat deposition. The present invention describes methods to detect the presence or absence of these single nucleotide polymorphisms of PLA2G1B, in particular G7328A and T9182G, and subsequently provide treatment that reduces fat deposition. This treatment may consist of an appetite suppressant, a lipase inhibitor, a phospholipase inhibitor, an exercise regimen, a dietary regimen, psychological counselling, psychotherapy or a psychotherapeutic. Accordingly, PLA2G1B is a target for reducing fat deposition and it can be used to treat both obesity and non-insulin dependent diabetes mellitus (NIDDM), as well as cardiovascular disorders such as hypertension. As such, it exhibits antidiabetic activity. This polypeptide sequence is the human PLA2G1B protein of the invention.

Alignment Scores: 1e-22 Length: 148
Pred. No.: 258.00 Matches: 55
Score: 59.8% Conservative: 6
Best Local Similarity: 53.9% Mismatches: 12
Query Match: 20.6% Indels: 29
DB: Gaps: 3

US-10-607-806-1-G7328_COPY_7000_7700 (1-701) x ADJ09984 (1-148)

QY 113 ATTTCTCTGAGCATCTTGGCGA-----AGT 139
DB 19 IleserProargAlaValTrpGlnPheArgLysMetIleLysCysValIleProGlySer 38
QY 140 CATCCCCACCTG-----TTCTCTGAGAGTGGCAGGTAGGGCTG 178
DB 39 AspProPheLeuGluTyrAsnAsnTyrGlyCysTyrCysGlyLeuGly-GlySerGly 58
QY 179 ACCTATTGCTCTGCATTTACTCTCTATCTCAGCTGTCCCTCCACTTTCCAGGTGCTGCCA 238
DB 58 rProValAsp-----GluleuAspLysCysCysGly 68
QY 239 GACACATGACAACTGCTATGACAGCCGCAAGAGCTGGACAGCTGTAAATTTCTGTGGA 298
DB 68 nThrHisAspAsnCysTyrAspGlnAlaLysLysLeuAspSerCysLysPheLeuLeuAs 88
QY 299 CAMMCCGTACACCCACACCTATTATCTATCTGCTGTGGCTGGCAATCACCTGTAGCAG 358
DB 88 pAsnProTyrThrHisThrTyrSerTyrSerCysSerGlySerAlaIleThrCysSerSe 108

QY 359 TAGG 362
DB 108 rLys 109

RESULT 9
AAB54059
ID AAB54059 standard; protein; 156 AA.
XX AAB54059;
XX 09-MAR-2001 (first entry)
XX Human pancreatic cancer antigen protein sequence SEQ ID NO:511.
XX Human; pancreas; pancreatic cancer; pancreatic cancer antigen; detection;
KW diagnosis; identification; cytostatic; neuroprotective; nootropic;
KW immunomodulatory; relaxant; contraceptive; gynaecological;
KW antiinflammatory; radiant; gene therapy; chromosome mapping;
KW linkage analysis; tissue identification; tissue typing; forensic; neural;
KW immune system; muscular; reproductive; gastrointestinal; pulmonary;
KW cardiovascular; renal; proliferative.
XX Homo sapiens.
XX WO200055320-A1.

XX PD 21-SEP-2000.
 XX XX 08-MAR-2000; 2000WO-US005989.
 XX PF 12-MAR-1999; 99US-0124270P.
 XX PR (HUMA-) HUMAN GENOME SCI INC.
 XX PA Rosen CA, Ruben SM;
 XX PI WPI; 2000-579444/54.
 XX DR N-PSDB; AAC98824.
 XX XX
 PT New nucleic acid that is a pancreatic cancer antigen for preventing,
 PT treating, or ameliorating a medical condition, particular pancreatic
 PT cancer, or for use in assays for diagnosing a pathological condition.
 XX
 PS Claim 11; Page 948; 1379pp; English.
 XX
 CC AAC98773 to AAC99231 encode the human pancreatic cancer associated
 CC proteins, called pancreatic cancer antigens, given in AAB54008 to
 CC AAB54466. The human pancreatic cancer antigens have cytostatic,
 CC neuroprotective, nontoxic, immunomodulatory, relaxant, contraceptive,
 CC gynaecological, cardiac and antiinflammatory activities, and can be used
 CC in gene therapy. The polynucleotide and proteins can be used for
 CC preventing, treating, or ameliorating a medical condition or in assays
 CC for diagnosing a pathological condition or a susceptibility to one in a
 CC subject. Binding partners to the proteins and the activity of the
 CC proteins can be identified. The pancreatic cancer antigens can be used to
 CC detect, treat or prevent pancreatic disorders, especially cancer.
 CC Agonists and antagonists to the antigens can be screened for. The
 CC pancreatic cancer antigen polynucleotides can be used to design nucleic
 CC acid hybridisation probes that can be used in chromosome mapping, linkage
 CC analysis, tissue identification and/or typing and a variety of forensic
 CC and diagnostic methods. The proteins can be used to generate antibodies
 CC which are used to purify, detect and target the polypeptides, including
 CC both in vivo and in vitro diagnostic and therapeutic methods. The
 CC proteins can be used to treat or prevent neural, immune system, muscular,
 CC reproductive, gastrointestinal, pulmonary, cardiovascular, renal or
 CC proliferative disorders. AAC99232 to AAC99240 and AAB54467 represent
 CC sequences used in the exemplification of the present invention
 XX
 SQ Sequence 156 AA;
 Alignment Scores:
 Pred. No.: 1.02e-22 Length: 156
 Score: 258.00 Matches: 55
 Percent Similarity: 59.8% Conservativity: 6
 Best Local Similarity: 53.9% Mismatches: 12
 Query Match: 20.6% Indels: 29
 DB: 3 Gaps: 3
 US-10-607-806-1-G7328_COPY_7000_7700 (1-701) x AAB54059 (1-156)
 QY 113 ATTCTCTCTCAAGCATCTTGGCGA-----AGT 139
 DB 27 IleserProArgalaValTrpGlnPheArgLysMetileLysCysValIleProGlySer 46
 QY 140 CATCCCCACCTG-----TTCTGTAGAGTGGCGAGGTGAGGCGTG 178
 DB 47 AspProPheLeuGluTyrAsnAsnTyrGlyCysTyrCysGlyLeuGly-GlySerGly 66
 QY 179 ACCTATTGTCTGCACTTACTCTCTAGCTGTCCCTCCCACTTCCAGGTGCTGCCA 238
 DB 66 rProValAsp-----GluLeuAspLysCysCysGly 76
 QY 239 GACACATGACAACTGTATGACAGCGCAAGAGCTGACAGCTGTAAATTTCTGTGGA 298
 DB 76 nThrHisAspAsnCysTyrAspGlnAlaLysLysLeuAspSerCysLysPheLeuAs 96
 QY 299 CMMCCGTACACCCACCACTTATTCATCTCTGCTCTGCTCGGCAATCACCTGTAGCAG 358

Db 96 pAsnProTyrThrHisThrTyrSerTyrSerCysSerGlySerAlaIleThrCysSerSe 116
 QY 359-TAGG 362
 Db 116 rLys 117
 RESULT 10
 AABM83696
 ID AABM83696 standard; protein; 194 AA.
 XX AABM83696;
 XX 18-NOV-2004 (first entry)
 XX Human diagnostic and therapeutic pprotein SEQ ID NO:3945.
 XX gene therapy; human diagnostic and therapeutic polynucleotide; dithp.
 XX Homo sapiens.
 XX WO2004023973-A2.
 XX 25-MAR-2004.
 XX 12-SEP-2003; 2003WO-US028227.
 XX 12-SEP-2002; 2002US-0410259P.
 XX 12-SEP-2002; 2002US-0410260P.
 XX (INCY-) INCYTE CORP.
 XX Schmidt JP, Wright RJ, Bruns CM, Marjanovic MM, Shen F;
 XX Harthshorne TA, Suchorolski MT, Altus CM, Pitts SJ, Elder LV;
 XX Hooney EM, Deleagane AM, Pancer SR, Banville SC, Reddy TP;
 XX Stevens KA, Blanchard JL, Panzer SR, Rioux P, Shen EJ, Wu MC, Gerstein EH;
 XX Peralta CH, Anderson SB, Rioux P, Shen EJ, Wu MC, Stuve LL;
 XX Lagace RE, Spiro PA, Stewart EA, Wingrove J, Vitt UA, Kirton ES;
 XX Xu Y, Kwong M, Policky JL, Hurwitz BL, Ma Y, Jackson JL, Gietzen D;
 XX Patursky S, Shi X, Suarez CJ;
 XX WPI; 2004-329368/30.
 XX N-PSDB; ACN42348.
 XX
 PT New diagnostic and therapeutic polynucleotides and polypeptides, useful
 PT in diagnosing a condition, disease or disorder associated with human
 PT molecules, e.g. autoimmune or inflammatory disorders, in gene therapy or
 PT in gene mapping.
 Claim 27; Page: 190pp; English.
 The invention relates to novel diagnostic and therapeutic polynucleotides
 selected from one of the 2722 sequences defined in the specification. A
 polynucleotide of the invention may have a use in gene therapy. The human
 diagnostic and therapeutic polynucleotides (dithp) or polypeptides may be
 used to diagnose a particular condition, disease or disorder associated
 with human molecules, e.g. cell proliferative disorders,
 autoimmune/inflammatory disorder, developmental disorder, endocrine
 disorder, neurological disorders, gastrointestinal disorders, or
 infections caused by virus, bacteria, fungi or parasite. The dithp
 molecules may also be used in genetic mapping, in identifying individuals
 from minute biological samples, in detecting single nucleotide
 polymorphisms, as molecular weight markers, and for somatic or germline
 gene therapy. The present sequence represents a dithp protein of the
 invention. Note: The sequence data for this patent is not represented in
 the printed specification, but was obtained in electronic format directly
 from WIPO at www.wipo.int/pct/en/sequences/listing.htm
 SQ Sequence 194 AA;
 Alignment Scores:
 Pred. No.: 1.11e-22 Length: 194
 Score: 258.00 Matches: 55
 Percent Similarity: 59.8% Conservativity: 6

US-10-607-806-1-G7328_COPY_7000_7700 (1-701) x AAB54421 (1-79)

QY 222 CTTTCCAGGTGCTGCCAGACACATGACAACTGCTAYGACAGGCCAAGAGCTGGACAGC 281
 Db 29 LeuAspLysCysCysGlnThrHisAspAsnCysTyrAspGlnAlaLys**LeuAspSer 48
 QY 282 TGTAAATTTCTGCTGGGACAMMCCGTACACACACCTATTTCATCTCGTCTGGCTCG 341
 Db 49 Cys**PheLeuLeuAspAsnProTyrThrHisThrTyrSerTyrSerCysSerGlySer 68

QY 342 GCAATCACCTGT 353
 Db 69 AlaIleThrCys 72

RESULT 14
 ABB08153
 ID ABB08153 standard; protein; 146 AA.

AC ABB08153;

DT 10-SEP-2002 (first entry)

DE Murine PLA2 enzyme mGI polypeptide.

XX Phospholipase A2; GXII PLA2; phosphatidylethanolamine; Th2; GV PLA2;
 KW transgenic; immunosuppressive; antiallergic; cytostatic; antimicrobial;
 KW antidiabetic; antirheumatic; antiarthritic; antiinflammatory; mouse;
 KW neuroprotective; cerebroprotective; antiinfertility; contraceptive;
 KW mGXII-1 PLA2; enzyme; mGI.

XX Mus musculus.

XX WO2002040655-A2.

XX 23-MAY-2002.

XX 06-NOV-2001; 2001WO-US04125.

XX 06-NOV-2000; 2000US-0246316P.

XX (HARD) HARVARD COLLEGE.

XX Ho I, Arm JP, Austen KF, Glimcher LH;

XX WPI; 2002-500219/53.

XX New group XII phospholipase A2 protein, useful for identifying modulators
 PT used for modulating prostaglandin production by Th2 cell and Th2 cell
 PT differentiation/activity, and treating allergy, cancer and type I
 PT diabetes.

XX Disclosure; Fig 1; 77pp; English.

XX The invention relates to an isolated group XII phospholipase A2 (GXII
 CC PLA2) protein or its biologically active portion, where the protein
 CC selectively hydrolyses arachidonic acid in sn-2 position of
 CC phosphatidylethanolamine. Methods of modulating Th2 cell differentiation
 CC activity by modulating either GXII PLA2 or GV PLA2 which is also
 CC preferentially expressed in T cells are provided. The GXII PLA2 proteins
 CC can be used to prepare anti-GXII PLA2 antibodies. The GXII PLA2 encoding
 CC nucleic acid molecules can be used to prepare non-human transgenic
 CC animals that contain cells carrying a transgene encoding GXII PLA2
 CC protein or a portion of GXII PLA2 protein. Prostaglandin production can
 CC be increased by stimulators of GXII PLA2 or GV PLA2 which is further
 CC useful for prolonging survival of the graft and thus has applications in
 CC bone marrow transplantation or solid organ transplantation, and for
 CC increasing production of Th2-promoting cytokines for commercial purposes.
 CC Modulating the type of T helper cell response mounted in the individual
 CC suffering from the disease condition e.g., inhibition of Th2 cell
 CC response using GXII PLA2 or GV PLA2 inhibitors is carried out for
 CC inhibiting prostaglandin production to thereby inhibit production of Th2-
 CC associated cytokines in: (a) allergic patients to downregulate production
 CC of pathogenic IgE antibodies; (b) cancer patients; and (c) subjects with

CC infectious diseases. Stimulation of Th2 cell response using GXII PLA2 or
 CC GV PLA2 stimulators is useful for treating autoimmune diseases associated
 CC with Th2-type dysfunction, such as type I diabetes, rheumatoid arthritis,
 CC etc. The methods provided are also useful in treating conditions such as
 CC inflammation, arthritis, multiple sclerosis, stroke, infertility, and
 CC also have contraceptive uses. The present sequence represents a mouse
 CC PLA2 enzyme mGI

SQ Sequence 146 AA;

Alignment Scores:
 Pred. No.: 3-32e-18 Length: 146
 Score: 222.00 Matches: 37
 Percent Similarity: 91.5% Conservative: 6
 Best Local Similarity: 78.7% Mismatches: 4
 Query Match: 17.7% Indels: 0
 DB: 5 Gaps: 0

US-10-607-806-1-G7328_COPY_7000_7700 (1-701) x ABB08153 (1-146)

QY 222 CTTTCCAGGTGCTGCCAGACACATGACAACTGCTAYGACAGGCCAAGAGCTGGACAGC 281

Db 63 LeuAspArgCysCysGlnThrHisAspHisCysTyrSerGlnAlaLysLysLeuGluSer 82

QY 282 TGTAAATTTCTGCTGGGACAMMCCGTACACACACCTATTTCATCTCGTCTGGCTCG 341

Db 83 CysLysPheLeuIleAspAsnProTyrThrAsnThrTyrSerTyrSerCysSerGlySer 102

QY 342 GCAATCACCTGTAGCAGTAGG 362

Db 103 GluIleThrCysSerAlaLys 109

RESULT 15

AAR63059

ID AAR63059 standard; protein; 124 AA.

XX AC AAR63059;

XX 25-MAR-2003 (revised)

DT 15-AUG-1995 (first entry)

XX Rat PLA2 type I.

XX RPLA2; RPLA2-8; RPLA2-10; phospholipase A2; PLA2.

XX Homo sapiens.

XX WO9502328-A1.

XX 26-JAN-1995.

PF 15-JUL-1994; 94WO-US007926.

XX 15-JUL-1993; 93US-00091941.

PR 26-JUL-1993; 93US-00097354.

XX (INDV) UNIV INDIANA FOUND.

XX (INCY-) INCYTE PHARM INC.

XX Tischfield JA, Seilhamer JJ;

XX WPI; 1995-067096/09.

XX Novel type III and IV low mol. wt. phospholipase A2 enzymes - from humans
 PT and rats, also nucleic acid sequences useful, e.g. for recombinant prodn.
 PT of enzymes, research into Batten's disease, etc.
 XX Disclosure; Fig 27; 160pp; English.
 XX Rat cDNAs (AA081136-37) encoding 2 novel PLA2s, RPLA2-8 (AAR63044) and
 CC RPLA2-10 (AAR63045), were isolated from brain and heart libraries,
 CC respectively. RPLA2-8 and RPLA2-10 were characterized as novel type III
 CC and IV PLA2s on the basis of their cysteine content compared with RPLA2

CC types I (AAR63059) and II (AAR63059). (Updated on 25-MAR-2003 to correct
 CC PN field.)
 XX
 SQ Sequence 124 AA;

Alignment Scores:
 Pred. No.: 1.77e-17 Length: 124
 Score: 216.00 Matches: 35
 Percent Similarity: 89.4% Conservative: 7
 Best Local Similarity: 74.5% Mismatches: 5
 Query Match: 17.2% Indels: 0
 DB: 2 Gaps: 0

US-10-607-806-1-G7328_COPY_7000_7700 (1-701) x AAR63059 (1-124)

Qy	222	CTTTCCAGGTGCTGCCAGACATGACAACTGCTAYGACCAAGCCAAAGAGCTGGACAGC	281
Db	41	LeuAspArgCysCysGlnThrHisAspHisCysTyrAsnGlnAlaLysLysLeuGluSer	60
Qy	282	TGTAATTTCTGCTGGACAMMCCGTACACCCACCTATTTCATCTGCTCTGGCTCG	341
Db	61	CysLysPheLeuIleAspAsnProTyrThrAsnThrTyrSerTyrLysCysSerGlyAsn	80
Qy	342	GCAATCACCTGTACAGTAGG	362
Db	81	ValIleThrCysSerAspLys	87

Search completed: February 8, 2006, 10:42:36
 Job time : 103.179 secs

THIS PAGE BLANK (USPTO)

GenCore version 5.1.7
Copyright (c) 1993 - 2006 Bioceleration Ltd.

OM nucleic - protein search, using frame_plus_n2p model

Run on: February 8, 2006, 10:44:07; Search time 4.78382 Seconds
(without alignments)
2819.838 Million cell updates/sec

Title: US-10-607-806-1-G7328_COPY_7000_7700
Perfect score: 1253
Sequence: 1 gtcgtgctactgtgtccag.....tgacgctcaacctctgag 701

Scoring table: BLOSUM62
Xgapop 10.0, Xgapext 0.5
Ygapop 10.0, Ygapext 0.5
Fgapop 6.0, Fgapext 7.0
Delop 6.0, Delext 7.0

Searched: 283416 seqs, 96216763 residues

Total number of hits satisfying chosen parameters: 566832

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Command line parameters:

-MODEL=frame+n2p.model -DEV=xlp
-Q=/abs/ABSSFE.spool/WARTIN10607806/runat_08022006_092020_27698/app_query.fasta_1
-DB=Pir -QMT=fastan -SUFFIX=n2p.rpr -MINMATCH=0.1 -LOOPCL=0 -LOOPEXT=0
-UNITS=bits -START=1 -END=1 -MATRIX=blosum62 -TRANS=human40.cdi -LIST=45
-DOCALIGN=200 -THR SCORE=pct -THR MAX=100 -THR MIN=0 -ALIGN=15 -MODE=LOCAL
-OUTFMT=ptc -NORM=ext -HEADSIZE=500 -MINLEN=0 -MAXLEN=2000000000 -HOST=abs02p
-USPR=WARTIN10607806 @CCGN 1.1 92 @runat_08022006_092020_27698 -NCPV=6 -ICPU=3
-NO MAP -NEG SCORES=0 -WAIT -DSPBLOCK=100 -LONGLOG -DEV_TIMEOUT=120
-WARN_TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5 -FGAPOP=6 -FGAPEXT=7
-YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database : Pir 80: *
1: Pir1: *
2: Pir2: *
3: Pir3: *
4: Pir4: *

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	258	20.6	148	1 PSBU	phospholipase A2
2	218	17.4	146	1 PSDG	phospholipase A2
3	216	17.2	144	2 JN0480	phospholipase A2
4	216	17.2	146	1 PSRT	phospholipase A2
5	209	16.7	145	1 PSBOA	phospholipase A2
6	205	16.4	146	1 PSPGA	phospholipase A2
7	202	16.1	124	1 PSPGA2	phospholipase A2
8	201.5	16.1	132	1 PSBOA	phospholipase A2
9	188	15.0	146	1 S34049	phospholipase A2
10	159	12.7	119	2 S00227	phospholipase A2
11	157	12.5	125	2 S65824	phospholipase A(2)
12	149.5	11.9	119	1 PSNJ3K	phospholipase A2
13	144.5	11.5	146	1 PSNJAF	phospholipase A2
14	143.5	11.5	119	1 PSNJ2K	phospholipase A2

ALIGNMENTS

RESULT 1

PSBU

phospholipase A2 (EC 3.1.1.4) IB precursor (validated) - human
N;Alternate names: pancreatic phospholipase A2; phosphatidylcholine 2-acylhydrolase
C;Species: Homo sapiens (man)
C;Date: 17-Mar-1987 #sequence_revision 31-Mar-1993 #text_change 16-Aug-2004
C;Accession: C25793; A25793; A00733; A21566
R;Seilhamer, J.J.; Randall, T.L.; Yamanaka, M.; Johnson, L.K.
DNA 5, 519-527, 1986

A;Title: Pancreatic phospholipase A2: isolation of the human gene and cDNAs from porcine
A;Reference number: A25793; MUID:87132925; PMID:3028739
A;Accession: C25793
A;Molecule type: DNA
A;Residues: 1-148 <SEI2>
A;Cross-references: UNIPROT:P04054; UNIPARC:UPI0000048FD9; GB:M22970
A;Accession: A25793
A;Molecule type: mRNA
A;Residues: 1-148 <SEI>
A;Cross-references: UNIPARC:UPI0000048FD9; GB:M21055
A;Experimental source: lung
R;Grataroli, R.; Dijkman, R.; Dutilh, C.E.; Van der Ouderaa, F.; De Haas, G.H.; Figarel
Eur. J. Biochem. 122, 111-117, 1982
A;Title: Studies on phospholipase A2 and its enzyme from human pancreatic juice. Cat
A;Reference number: A91114; MUID:82138816; PMID:7060561
A;Accession: A00733
A;Molecule type: protein
A;Residues: 16-22 <VER1>
A;Cross-references: UNIPARC:UPI000014C3BF
R;Verheij, H.M.; Westerman, J.; Sternby, B.; De Haas, G.H.
Biochim. Biophys. Acta 747, 93-99, 1983
A;Title: The complete primary structure of phospholipase A2 from human pancreas.
A;Reference number: A90651; MUID:83283533; PMID:6349696
A;Accession: A21566
A;Molecule type: protein
A;Residues: 23-143, 'YS', 146-147 <VER2>
A;Cross-references: UNIPARC:UPI0000172758
C;Genetics:
A;Gene: GDB:PLA2G1B; PLA2A; PLA2; P

A;Cross-references: GDB:120715; OMIM:172410
A;Map position: 12q23-12qter
A;Introns: 12/1; 65/2; 108/1
C;Function:

A;Description: catalyzes hydrolysis of 1,2-diacyl-sn-glycero-3-phosphocholine to 1-acyl-
A;Note: the reaction is strongly enhanced when the phospholipid is condensed into a micelle
C;Superfamily: Phospholipase A2
C;Keywords: calcium; carboxylic ester hydrolase; lipid degradation; lipid digestion; metabolism
F;1-15/Domain: signal sequence #status predicted <SIG>
F;16-22/Domain: activation peptide #status experimental <MAT>
F;23-148/Product: phospholipase A2 IB #status predicted
F;33-99,49-146,51-67,127,73-120,83-113,106-118/Disulfide bonds: #status predicted
F;50,52,54,71/Binding site: calcium (Tyr, Gly, Asp) #status predicted
F;70,121/Active site: His, Asp #status predicted

Alignment Scores:
Pred. No.: 2,99e-20 Length: 148
Score: 258.00 Matches: 55
Percent Similarity: 59.8% Conservative: 6
Best Local Similarity: 53.9% Mismatches: 12
Query Match: 20.6% Indels: 29
DB: 1 Gaps: 3

US-10-607-806-1-G7328_COPY_7000_7700 (1-701) x PSBU (1-148)

QY 113 ATTTCTCTGAGCATCTTGGCGA-----AGT 139
DB 19 ILESerProArgAlaValTrpGlnPheArguysMetilelysCysValIleProGlySer 38
QY 140 CATCCCCACCTG-----TTCTGAGAGTGGCAGGTGAGGCTG 178
DB 39 AspProPheLeuGluTyrAsnAsnTyrGlyCysTyrCysGlyLeuGly-GlySerGlyTh 58
QY 179 ACCTATTGCTGTCACCTTACTCTCTATCTCAGCTGTCCCTCCCACTTTCCAGGTGCTGCCA 238
DB 58 rProValAsp-----GluleuAspLysCysCysGI 68
QY 239 GACACATGACACTGCTAYGACAGGCCAAGAGCTGGACAGCTGTAAATTTCTGCTGGA 298
DB 68 nThrHisAspAsnCysTyrAspGlnAlaLysLysLeuAspSerCysLysPheLeuAs 88
QY 299 CMMCCGTACACCCACACCTATTCTGCTGTGCTGGCTCGGCAATCACTGTAGCAG 358
DB 88 pAsnProTyrThrHisThrThrSerTyrSerCysSerGlySerAlaIleThrCysSerSe 108
QY 359 TAGG 362
DB 108 rLys 109

RESULT 2

PSDG
phospholipase A2 (EC 3.1.1.4) precursor - dog
N;Alternate names: phosphatidylcholine 2-acylhydrolase
C;Species: Canis lupus familiaris (dog)
C;Date: 30-Sep-1987 #sequence revision 30-Sep-1987 #text_change 16-Aug-2004
C;Accession: S11316; A24392; J50006
R;Kerfelec, B.; LaForge, K.S.; Vasiloudes, P.; Puigserver, A.; Scheele, G.A.
Eur. J. Biochem. 190, 299-304, 1990
A;Title: Isolation and sequence of the canine pancreatic phospholipase A(2) gene.
A;Reference number: S11316; MUID:90306027; PMID:2142076
A;Accession: S11316
A;Molecule type: DNA
A;Residues: 1-146 <K2>
A;Cross-references: UNIPROT:P06596; UNIPARC:UPI0000013108C
R;Kerfelec, B.; LaForge, K.S.; Puigserver, A.; Scheele, G.
Pancreas 1, 430-437, 1986
A;Title: Primary structures of canine pancreatic lipase and phospholipase A2 messenger R
A;Reference number: A93751; MUID:87175472; PMID:3562437
A;Accession: A24392
A;Molecule type: mRNA
A;Residues: 1-146 <KER>
A;Cross-references: UNIPARC:UPI0000013108C; GB:M35301; NID:g164041; PIDN:AAA30883.1; PID:
R;Ohara, O.; Tamaki, M.; Nakamura, E.; Tsuruta, Y.; Fujii, Y.; Shin, M.; Teraoka, H.; Oh

J. Biochem. 99, 733-739, 1986
A;Title: Dog and rat pancreatic phospholipases A2: complete amino acid sequences deduced
A;Reference number: A92008; MUID:86223862; PMID:3754861
A;Accession: J50006
A;Molecule type: mRNA
A;Residues: 1-146 <OHA>
A;Cross-references: UNIPARC:UPI000013108C; GB:D00035; NID:g217659; PIDN:BAA000C

C;Function:
A;Description: catalyzes hydrolysis of 1,2-diacyl-sn-glycero-3-phosphocholine to 1-acyl-
A;Note: the reaction is strongly enhanced when the phospholipid is condensed into a micelle
C;Superfamily: Phospholipase A2
C;Keywords: calcium; carboxylic ester hydrolase; lipid degradation; lipid digestion; metabolism
F;1-15/Domain: signal sequence #status predicted <SIG>
F;16-22/Domain: activation peptide #status predicted <MAT>
F;23-146/Product: phospholipase A2 #status predicted
F;33-99,49-146,51-67,127,73-120,83-113,106-118/Disulfide bonds: #status predicted
F;50,52,54,71/Binding site: calcium (Tyr, Gly, Asp) #status predicted
F;70,121/Active site: His, Asp #status predicted

Alignment Scores:
Pred. No.: 8,9e-16 Length: 146
Score: 218.00 Matches: 48
Percent Similarity: 54.9% Conservative: 8
Best Local Similarity: 47.1% Mismatches: 17
Query Match: 17.4% Indels: 29
DB: 1 Gaps: 3

US-10-607-806-1-G7328_COPY_7000_7700 (1-701) x PSDG (1-146)

QY 113 ATTTCTCTGAGCATCTTGGCGA-----AGT 139
DB 19 ILESerProArgAlaValTrpGlnPheArgAsnMetilelysCysThrIleProGluSer 38
QY 140 CATCCC-----CACCTGTTCTCAGAGTGGCAGGTGAGGCTG 178
DB 39 AspProLeuLysAspTyrAsnAspTyrGlyCysTyrCysGlyLeuGly-GlySerGlyTh 58
QY 179 ACCTATTGCTGTCACCTTACTCTATCTCAGCTGTCCCTCCCACTTTCCAGGTGCTGCCA 238
DB 58 rProValAsp-----GluleuAspLysCysCysGI 68
QY 239 GACACATGACACTGCTAYGACAGGCCAAGAGCTGGACAGCTGTAAATTTCTGCTGGA 298
DB 68 nThrHisAspHisCysTyrSerGluAlaLysLysLeuAspSerCysLysPheLeuAs 88
QY 299 CMMCCGTACACCCACACCTATTCTGCTGTGCTGGCTCGGCAATCACTGTAGCAG 358
DB 88 pAsnProTyrThrLysIleTyrSerTyrSerCysSerGlySerGluIleThrCysSerSe 108
QY 359 TAGG 362
DB 108 rLys 109

RESULT 3

JN0480
phospholipase A2 (EC 3.1.1.4) precursor, pancreatic type - rabbit
N;Alternate names: phosphatidylcholine 2-acylhydrolase
C;Species: Oryctolagus cuniculus (domestic rabbit)
C;Date: 30-Sep-1993 #sequence revision 30-Sep-1993 #text_change 05-Oct-2004
C;Accession: JN0480
R;Kumar, V.B.
Biochem. Biophys. Res. Commun. 192, 683-692, 1993
A;Title: Cloning and expression of rabbit pancreatic phospholipase A2.
A;Reference number: JN0480; MUID:93249443; PMID:7916604
A;Accession: JN0480
A;Molecule type: DNA
A;Residues: 1-144 <KUM>
A;Cross-references: UNIPROT:Q7M334; UNIPARC:UPI000017586E; GB:L11899
C;Superfamily: Phospholipase A2
C;Keywords: calcium; carboxylic ester hydrolase; lipid digestion; zymogen
F;1-15/Domain: signal sequence #status predicted <SIG>
F;16-22/Domain: activation peptide #status predicted <PRO>
F;23-144/Product: phospholipase A2, pancreatic type #status predicted <MAT>

tatus predicted
icted

ACAGC 281

luSer 82

lyAsn 102

-acylhydrolase

9-2004

e A2.

0; NID:9634; PIDN:(

sheep pancreas.

de bonds

R;Dijkstra, B.W.; Drenth, J.; Kalk, K.H.
 Nature 289, 604-606, 1981
 A;Title: Active site and catalytic mechanism of phospholipase A2.
 A;Reference number: A93243; MUID:81123082; PMID:7464926
 C;Contents: annotation; X-ray crystallography, 1.7 angstroms; active site; catalytic mechanism
 R;Dijkstra, B.W.; van Nes, G.J.H.; Kalk, K.H.; Brandenburg, N.P.; Hol, W.G.J.; Drenth, J.
 submitted to the Brookhaven Protein Data Bank, June 1981
 A;Reference number: A50406; PDB:2BP2
 A;Contents: annotation; X-ray crystallography, 3.0 angstroms, residues 23-143, 'N', 145
 A;Note: pancreas, phospholipase A2 precursor
 R;Dijkstra, B.W.; van Nes, G.J.H.; Kalk, K.H.; Brandenburg, N.P.; Hol, W.G.J.; Drenth, J.
 Acta Crystallogr. B38, 793-799, 1982
 A;Title: The structure of bovine pancreatic phospholipase A2 at 3.0 angstroms resolution
 A;Reference number: A58516
 A;Contents: annotation; X-ray crystallography, 3.0 angstroms
 R;Sundaralingam, M.
 submitted to the Brookhaven Protein Data Bank, January 1992
 A;Reference number: A51455; PDB:2BPP
 A;Contents: annotation; X-ray crystallography, 1.8 angstroms, residues 23-143, 'N', 145
 A;Note: recombinant form expressed in *Escherichia coli*
 R;Noel, J.P.; Bingman, C.A.; Deng, T.; Dupreux, C.M.; Hamilton, K.J.; Jiang, R.T.; Kwak
 Biochemistry 30, 11801-11811, 1991
 A;Title: Phospholipase A2 engineering. X-ray structural and functional evidence for the
 A;Reference number: A58513; MUID:92089090; PMID:1751497
 A;Contents: annotation; X-ray crystallography, 1.8 angstroms
 C;Function:
 A;Description: catalyzes hydrolysis of 1,2-diacyl-sn-glycero-3-phosphocholine to 1-acyl-
 A;Note: the reaction is strongly enhanced when the phospholipid is condensed into a micelle
 C;Superfamily: Phospholipase A2
 F;1-15/Domain: signal sequence #status predicted <SIG>
 F;16-145/Product: phospholipase A2 #status experimental <PRO>
 F;16-22/Domain: activation peptide #status experimental <APT>
 F;23-99/Modified site: pyrrolidone carboxylic acid (Gln) (in precursor form) #link PRO #sta
 F;33-99, 49-145, 51-67, 66-127, 73-120, 83-113, 106-118/disulfide bonds: #status experimental
 F;50, 52, 54, 71/Binding site: calcium (Tyr, Gly, Asp) #status experimental
 F;70, 121/Active site: His, Asp #status experimental

Alignment Scores:

Pred. No.:	9.04e-15	Length:	145
Score:	209.00	Matches:	36
Percent Similarity:	84.8%	Conservative:	3
Best Local Similarity:	78.3%	Mismatches:	7
Query Match:	16.7%	Indels:	0
DB:	1	Gaps:	0

US-10-607-806-1-G7328_COPY_7000_7700 (1-701) x PSBOA (1-145)
 Qy 222 CTTTCCAGGTCCTCCAGACATGACAACTGCTAYGACGAGCCAAAGACTGGACAGC 281
 Db 63 LeuAspArgCysGlnThrHisAspAsnCysTyrIysGlnAlaLysLeuAspSer 82
 Qy 282 TGTAAATTCCTGGGACMMCCGTACACCCACCTATTCTGCTCTGCTGGCTCG 341
 Db 83 CysLysValLeuValAspAsnProTyrThrAsnAsnTyrSerTyrSerCysSerAsnAsn 102
 Qy 342 GCATTCACCTTAGCAGT 359
 Db 103 GluileThrCysSerSer 108

RESULT 6
 PSPGA
 phospholipase A2 (EC 3.1.1.4) major precursor [validated] - pig
 N;Alternate names: pancreatic phospholipase A2; phosphatidylcholine 2-acylhydrolase
 C;Species: Sus scrofa domestica (domestic pig)
 C;Date: 24-Apr-1984 #sequence revision 30-Jun-1993 #text_change 16-Aug-2004
 C;Accession: B25793; S30750; A60264; A00734
 R;Seilhauer, J.J.; Randall, T.L.; Yamanaka, M.; Johnson, L.K.
 DNA 5, 519-527, 1986
 A;Title: Pancreatic phospholipase A2: isolation of the human gene and cDNAs from porcine
 A;Reference number: A25793; MUID:87132925; PMID:3028739
 A;Accession: B25793

A;Molecule type: mRNA
 A;Residues: 1-146 <SEI>
 A;Cross-references: UNIPROT:P00592; UNIPARC:UPI0000113109C; EMBL:M21055; NID:G164611; PID:
 R;de Guss, P.; van den Bergh, C.J.; Kuipers, O.; Verheij, H.M.; Hoekstra, W.P.M.; de Haas
 Nucleic Acids Res. 15, 3743-3759, 1987
 A;Title: Expression of porcine pancreatic phospholipase A(2). Generation of active enzyme
 A;Reference number: S30750; MUID:87231066; PMID:3295782
 A;Accession: S30750
 A;Molecule type: mRNA
 A;Residues: 1-146 <DEG>
 A;Cross-references: UNIPARC:UPI000013109C; EMBL:Y00146; NID:G2051; PID:CAA68341.1; PID:
 R;Kuipers, O.P.; van den Bergh, C.J.; Verheij, H.M.; de Haas, G.H.
 Adv. Exp. Med. Biol. 279, 65-84, 1990
 A;Title: Probing the mechanism of pancreatic phospholipase A2 with the aid of recombinant
 A;Reference number: A60264; MUID:91263889; PMID:2096701
 A;Accession: A60264
 A;Status: not compared with conceptual translation
 A;Molecule type: mRNA
 A;Residues: 1-146 <KUI>
 A;Cross-references: UNIPARC:UPI000013109C
 R;Puijk, W.C.; Verheij, H.M.; de Haas, G.H.
 Biochim. Biophys. Acta 492, 254-259, 1977
 A;Title: The primary structure of phospholipase A-2 from porcine pancreas. A reinvestigat
 A;Reference number: A90620; MUID:77222092; PMID:884127
 A;Contents: revision
 A;Accession: A00734
 A;Molecule type: protein
 A;Residues: 16-146 <PUI>
 A;Cross-references: UNIPARC:UPI0000111B3
 R;de Haas, G.H.; Slotboom, A.J.; Bensen, P.P.M.; Nieuwenhuizen, W.; van Deenen, L.L.M.; N
 Biochim. Biophys. Acta 221, 54-61, 1970
 A;Title: Studies on phospholipase A and its zymogen from porcine pancreas. II. The assign
 A;Reference number: A90579; MUID:71014236; PMID:4919729
 A;Contents: annotation; disulfide bonds
 R;Dijkstra, B.W.; Renetseder, R.; Kalk, K.H.; Hol, W.G.J.; Drenth, J.
 submitted to the Brookhaven Protein Data Bank, June 1983
 A;Reference number: A50305; PDB:1P2P
 A;Contents: annotation; X-ray crystallography, 2.6 angstroms
 R;Dijkstra, B.W.; Renetseder, R.; Kalk, K.H.; Hol, W.G.J.; Drenth, J.
 J. Mol. Biol. 168, 163-179, 1983
 A;Title: Structure of porcine pancreatic phospholipase A-2 at 2.6 angstrom resolution and
 A;Reference number: A92899; MUID:83268704; PMID:6876174
 A;Contents: annotation; X-ray crystallography, 2.6 angstroms; active and binding sites
 R;van den Bergh, C.J.; Slotboom, A.J.; Verheij, H.M.; de Haas, G.H.
 J. Cell. Biochem. 39, 379-390, 1989
 A;Title: The role of Asp-49 and other conserved amino acids in phospholipases A2 and the
 A;Reference number: A60719; MUID:89255682; PMID:2722967
 A;Contents: annotation
 R;Finzel, B.C.; Ohlendorf, D.H.; Weber, P.C.; Salemme, F.R.
 submitted to the Brookhaven Protein Data Bank, October 1991
 A;Reference number: A50680; PDB:4P2P
 A;Contents: annotation; X-ray crystallography, 2.4 angstroms, residues 23-146
 R;Finzel, B.C.; Ohlendorf, D.H.; Weber, P.C.; Salemme, F.R.
 Acta Crystallogr. B 47, 558-559, 1991
 A;Title: An independent crystallographic refinement of porcine phospholipase A2 at 2.4 a
 A;Reference number: A58515; MUID:92029796; PMID:1930837
 A;Contents: annotation; X-ray crystallography, 2.4 angstroms
 C;Function:
 A;Description: catalyzes hydrolysis of 1,2-diacyl-sn-glycero-3-phosphocholine to 1-acyl-
 A;Note: the reaction is strongly enhanced when the phospholipid is condensed into a micelle
 C;Superfamily: Phospholipase A2
 F;1-15/Domain: signal sequence #status predicted <SIG>
 F;23-146/Product: phospholipase A2 major #status experimental <APT>
 F;16/Modified site: pyrrolidone carboxylic acid (Gln) (in mature form) #status experimen
 F;33-99, 49-146, 51-67, 66-127, 73-120, 83-113, 106-118/disulfide bonds: #status experimental
 F;50, 52, 54, 71/Binding site: calcium, high affinity (Tyr, Gly, Asp) #status experime
 F;93, 94, 121/Binding site: calcium, low affinity (Glu, Ser, Glu) #status predicted

Alignment Scores:

Pred. No.:	2,53e-14	Length:	146
Score:	205.00	Matches:	44
Percent Similarity:	53.9%	Conservative:	11
Best Local Similarity:	43.1%	Mismatches:	18
Query Match:	16.4%	Indels:	29
DB:	1	Gaps:	3

US-10-607-806-1-G7328 COPY 7000 7700 (1-701) x PSPGA (1-146)

QY	113	ATTCTCTCGAAGCACTCTGGCGA-----AGT	139
Db	19	IIeSerSerArgAlaLeuTriGlnPheArgSerMetIleLysCysAlaIleProGlySer	38
QY	140	CATCCCCACCTG-----TTCTGAGAGTGGCAGGTGAGGGCTG	178
Db	39	HisProLeuMetAspPheAsnAsnTyrGlyCysTyrCysGlyLeuGly-GlySerGlyTh	58
QY	179	ACCTATTGCTCTGCACCTTACTCCCTATCTCAGCTGTCCCTCCCACTTTCAGGTCGTGCCA	238
Db	58	rProValAsp-----GlueuAspArgCysCyl	68
QY	239	GACACATGACAACTGCTAYGACCAGGCCAAGCTGCAGCTGTAAATTTCTGCTGCA	298
Db	68	uThrHisAspAsnCysTyrArgAspAlaLysAsnLeuAspSerCysLysPheLeuValAs	88
QY	299	CAMMCCGTACACCCACACTATTATCTCGCTCTGGCTCGCGCAATCACCTGTAGCAG	358
Db	88	paenProTyrThrGluSerTyrSerTyrSerCysSerAsnThrGluIleThrCysAsnSe	108
QY	359	TAGG	362
Db	108	rLys	109

RESIST. 7

N/Phospholipase A2 (EC 3.1.1.4) minor - pig (tentative sequence)
 N/Alternate names: pancreatic phospholipase A2 minor isozyme; phosphatidylcholine 2-acyl
 C/Species: *Sus scrofa domestica* (domestic pig)
 C/Date: 17-Mar-1987 #sequence_revision 17-Mar-1987 #text_change 16-Aug-2004
 C/Accession: A00735
 R/Fuijlk, W.C.; Verheij, H.M.; Wietzes, P.; de Haas, G.H.
 Biochim. Biophys. Acta 580, 411-415, 1979
 A/Title: The amino acid sequence of the phospholipase A-2 isoenzyme from porcine pancrea
 A/Reference number: A00735; MUID:80088382; PMID:518908

Alignment Scores:		
Pred. No.:	5.55e-14	Length:
Score:	202.00	Matches:
Percent Similarity:	85.1%	Conservative:
Best Local Similarity:	70.2%	Mismatches:
Query Match:	66.1%	Indels:
DB:	1	Gaps:
		124
		33
		7
		0
		0

US-10-607-806-1-G7328 COPY 7000 7700 (1-701) x PSPGA2 (1-124)

Qy 222 CTTTCCAGGTGCTGCCAGACATGCACAACACTGCTAYGACCAGGCCAAGAAGCTGGACAGC 281
||| |||||||:::||||| ||||| ||||| ||||| |||||
Db 41 LeuAspArgCysvsgluthrHisaspAsnCystvYrAspAlalysAsnleuAspSer 60

282 TGTAAATTTCTGCTGGACAMCCCTACACCCACACCTATTATATCTGCTCTGCTGCTG 341 QY
61 CysLysSheuLeuValAspAsnProTyrThrAsnSerTyrSerTyrSerCysSerAsnThr 80 Db
342 GCAATCACCTGTAGCAGTAGG 362 QY
81 GluIleThrCysAsnSerLys 87 Db

RESULT 8

PSHOA

phospholipase A2 (EC 3.1.1.4) precursor - horse (tentative sequence)
N/Alternate names: pancreatic phospholipase A2; phosphatidylcholine 2-acylhydrolase
C/Species: Equus caballus (domestic horse)
C/Date: 24-Apr-1984 #sequence revision 03-Aug-1984 #text_change 16-Aug-2004
C/Accession: A90615; A92198; A90651; A00737
R/Evenberg, A.; Meyer, H.; Verheij, H.M.; de Haas, G.H.
Biochim. Biophys. Acta 491, 265-274, 1977
A/Title: Isolation and properties of phospholipase A-2 and phospholipase A-2 from horse pancreas.
A/Reference number: A90615; MUID:77134902; PMID:849461
A/Accession: A90615
A/Molecule type: protein
A/Residues: 1-7 <EVE>
A/Cross-references: UNIPROT:P00594; UNIPARC:UPI000017275B
A/Note: a second form, lacking residues 1 and 2, was also found
R/Evenberg, A.; Meyer, H.; Gaatstra, W.; Verheij, H.M.; de Haas, G.H.
J. Biol. Chem. 252, 1189-1196, 1977
A/Title: Amino acid sequence of phospholipase A-2 from horse pancreas.
A/Reference number: A92198; MUID:77118587; PMID:838712
A/Contents: active site
A/Accession: A92198
A/Molecule type: protein
A/Residues: 8-131 <EV2>
A/Cross-references: UNIPARC:UPI000017275C
R/Verheij, H.M.; Westerman, J.; Sternby, B.; De Haas, G.H.
Biochim. Biophys. Acta 747, 93-99, 1983
A/Title: The complete primary structure of phospholipase A2 from human pancreas.
A/Reference number: A90651; MUID:83283533; PMID:6349696

Alignment Scores:	6.28e-14	Length:	132
Pred. No.:	201.50	Matches:	48
Score:	43.5%	Conservative:	19
Percent Similarity:	31.2%	Mismatches:	31
Best Local Similarity:	16.1%	Indels:	56
Query Match:		Gaps:	6
DB:	1		

US-10-607-806-1-G7328 COPY 7000 7700 (1-701) X PSHOA (1-132)

113 ATTTCTCTCTGAAGCATCTTGGCGA-----AGT 139
 4 flleSerProArgAlaValTrpGlnPheArgSerMetIleGlnCysThrIleProAsnSer 23
 140 CATCCCCACCTG-----TTCTCTGAGAGTGGCGAGTGTAGGCGTG 178
 24 LysProTyrLeuGluPheAsnAspTyr-GlyCysTyrGlyLeuGly-GlySerGly 43
 179 ACCTATTGCTCTGCACCTTACTCCTTATCTAGCTGTCCCTCCCACTTTCCAGGTGTGCCA 238


```

Db      43 rProValasp-----|||-----GluleuAspAlaCysCysGI 53
Qy      239 GACACATGACAACTGTAYGACCAAGGCAAGAGCTGGACAGCTGTAAATTTCTGTGGGA 298
Db      53 nValHisAspAsnCysTyrThrGlnAlaLysGluLeuSerSerCysArgPheLeuValAs 73
Qy      299 CAMMCGGTACACCAACCTATTTCATCTCGTGTCTCGGCTCGGCATCACCCTGTAGCAG 358
Db      73 pAsnProTyrThrGluSerTyrLysPheSerCysSerGlyThrGluValThrCysSerAs 93
Qy      359 TAGCTTTATCCCTTCCTTGCATTAATCTAGTGGTCTCAGTAGCGCGGGGGA 418
Db      93 pLys-----|||-----As 95
Qy      419 TAATAGT-----AACACAGCCATGATTTAGTGTAAATTTCTTGGT 460
Db      95 nAsnAlaCysGluAlaPheIleCysAsnCysAspArgAsnAlaAlaIleCysPhe----- 113
Qy      461 TCTGGGAGTGTCTCCTTTAATCCTCAGAACACACATATG 500
Db      114 ----SerLysAlaProTyrAsnProGluAsnLysAsnLeu 125

RESULT 9
S34049
phospholipase A2 (EC 3.1.1.4) precursor, gastric and pancreatic [validated] - guinea pig
N;Alternate names: pancreatic phospholipase A2; phosphatidylcholine 2-acylhydrolase
C;Species: Cavia porcellus (guinea pig)
C;Date: 18-Aug-2000 #sequence_revision 18-Aug-2000 #text_change 16-Aug-2004
C;Accession: S34049; S34046; S34048; S34047
R;Ying, Z.; Tojo, H.; Nonaka, Y.; Okamoto, M.
Eur. J. Biochem. 215, 91-97, 1993
A;Title: Cloning and expression of phospholipase A(2) from guinea pig gastric mucosa, id
A;Reference number: S34049; MUID:93345504; PMID:8344290
A;Accession: S34049
A;Molecule type: mRNA
A;Residues: 1-146 <YIN>
A;Cross-references: UNIPROT:P43434; UNIPARC:UPI000013108D; GB:D00740; NID:G398128; PIDN:
R;Tojo, H.; Ying, Z.; Okamoto, M.
Eur. J. Biochem. 215, 81-90, 1993
A;Title: Purification and characterization of guinea pig gastric phospholipase A(2) of b
A;Reference number: S34046; MUID:93345502; PMID:8344288
A;Accession: S34046
A;Status: preliminary
A;Molecule type: protein
A;Residues: 23-63 <TOU>
A;Cross-references: UNIPARC:UPI000000E7305
A;Experimental source: gastric mucosa
A;Accession: S34048
A;Status: preliminary
A;Molecule type: protein
A;Residues: 23-24, 'X', 26-32, 'X', 34 <T02>
A;Cross-references: UNIPARC:UPI000017275D
A;Experimental source: pancreas
C;Comment: In this species phospholipase activity is high in gastric mucosa and jueic an
C;Function:
A;Description: catalyzes hydrolysis of 1,2-diacyl-sn-glycero-3-phosphocholine to 1-acyl-
A;Note: the reaction is strongly enhanced when the phospholipid is condensed into a micel
A;Superfamily: Phospholipase A2
C;Keywords: calcium; carboxylic ester hydrolase; gastric juice; lipid degradation; lipid
F;16-22/Domain: activation peptide #status predicted <ACT>
F;23-146/Product: phospholipase A2 #status experimental <NAT>
F;33-99, 49-146, 51-67, 66-127, 73-120, 83-113, 106-118/Disulfide bonds: #status predicted
F;50, 52, 54, 71/Binding site: calcium (Phe, Gly, Asp) #status predicted
F;70, 121/Active site: His, Asp #status predicted

Alignment Scores:
Pred. No.: 2, 02e-12 Length: 146
Score: 188.00 Matches: 41
Percent Similarity: 53.2% Conservative: 17
Best Local Similarity: 37.6% Mismatches: 22
Query Match: 15.0% Indels: 29

```

```

DB:      1      3      Gaps:
US-10-607-806-1-G7328_COPY_7000_7700 (1-701) x S34049 (1-146)
Qy      92 ATTTTTCAGCCATAGGATCATTTCTCTGAGCAATCTTGGCGA----- 136
Db      12 ValSerAlaAlaHisIleThrProArgAlaLeuTrpGlnPheArgAspMetIle 31
Qy      137 -----AGTCATCCCAC-----CTGTTCCTG 157
Db      32 LysCysAlaIleProGlySerArgProTyrSerGluTyrAsnAsnTyrGlyCysPheCys 51
Qy      158 AGAGTGGGAGGTGAGGCTGACCTATTGCTCTGCACCTTACTCTATCTCAGCTGTCCT 217
Db      52 GlyLeuGly-GlySerGlyThrProValasp----- 61
Qy      218 CCACATTTCCAGGTGTGCGACACACATGACAACTGCTAYGACCAGGCCAAGAGCTGA 277
Db      62 -GluLeuAspArgCysCysGluIleHisAspAlaCysTyrThrGlnAlaLysHisLeuGI 81
Qy      278 CAGCTGTAAATTTCTGCTGGACAMMCCGTACACCCACACTATTTCATCTCGTCTCTGG 337
Db      81 uSerCysLysSerValIleAspAsnProTyrThrAsnSerTyrSerPheSerCysSerGI 101
Qy      338 CTGGCAATCACCTGTAGCAGTAGG 362
Db      101 YThrAsnIleIleCysSerSerLys 109

RESULT 10
S00227
phospholipase A2 (EC 3.1.1.4) - shield snake
C;Species: Aspidelaps scutatus (shield snake)
C;Date: 31-Dec-1988 #sequence_revision 31-Dec-1988 #text_change 05-Oct-2004
C;Accession: S00227
R;Joubert, F.J.
Biol. Chem. Hoppe-Seyler 368, 1597-1602, 1987
A;Title: Purification, some properties of two phospholipases A2 (CM-I and CM-II) and the
A;Reference number: S00227; MUID:88163080; PMID:3442602
A;Accession: S00227
A;Molecule type: protein
A;Residues: 1-119 <YOU>
A;Cross-references: UNIPROT:P07037; UNIPARC:UPI00001310AB
A;Note: the sequence from Fig. 4 is inconsistent with that from Fig. 3 in having 61-Leu
C;Comment: Phospholipase A2 hydrolyzes the 2-acyl groups in 3-sn-phosphoglycerides.
C;Superfamily: Phospholipase A2
C;Keywords: calcium; carboxylic ester hydrolase; homodimer; lipid degradation; venom
F;1-119/Product: phospholipase A2 #status experimental <NAT>
F;11-71, 26-118, 28-44, 43-99, 50-92, 60-85, 78-90/Disulfide bonds: #status predicted
F;47, 93/Active site: His, Asp #status predicted

Alignment Scores:
Pred. No.: 3, 58e-09 Length: 119
Score: 159.00 Matches: 38
Percent Similarity: 41.1% Conservative: 6
Best Local Similarity: 35.5% Mismatches: 25
Query Match: 12.7% Indels: 38
Db:      2      Gaps: 5
US-10-607-806-1-G7328_COPY_7000_7700 (1-701) x S00227 (1-119)
Qy      66 AATAAATATATGATAAAAGGACTTATA-----TTT 95
Db      1 AsnLeuTyrGlnPheLysAsnMetIleGlnCysThrValProAsnArgSerTyrTrpHis 20
Qy      96 TTTCAGCCATAGATCATTTCTCTGAGCATCTTGGCGAAGTCATCCCCACCTGTCC 155
Db      21 PheAlaAspTyrGlyCysPhe-----CysGly 29
Qy      156 TGAGAGTGGGCGAGGTGAGGCTGACCTATTGCTGCTGACCTTACTCTATCTCAGCTGTCC 215
Db      30 -----TyrGlyGlySerGlyThrProValasp----- 38
Qy      216 CTCCCACTTTCCAGGTGCTGCCAGACACATGACAACTGTGTAYGACCAGGCCAAGAAGCTG 275

```

Db 39 ---GluLeuAspArgCysGlnThrHisAspAsnCysTyrSerGluAlaGluLysLeu 57
QY 276 GACAGCTGTAATTTCTGCTGGACAMMCGGTACACCCACCTATTTCATCTGCTGCTCT 335
Db 58 SerGlyCysLys-----ProfTyrlleLysThrTyrSerTyrAspCysSer 72
QY 336 GGCTCGCAATCACTCTAGC 356
Db 73 GlnGlyLysLeuThrCysSer 79
RESULT 11
S65624
phospholipase A(2) - common tiger snake
C:Species: Notechis scutatus scutatus (common tiger snake, mainland tiger snake)
C:Date: 19-Mar-1997 #sequence_revision 25-Apr-1997 #text_change 05-Oct-2004
C:Accession: S65624
R:Francis, B.; Coffield, J.A.; Simpson, L.L.; Kaiser, I.I.
Arch. Biochem. Biophys. 318, 481-488, 1995
A:Title: Amino acid sequence of a new type of toxic phospholipase A(2) from the venom of
A:Reference number: S65624; MUID:95251401; PMID:7733680
A:Accession: S65624
A>Status: preliminary
A:Molecule type: protein
A:Residues: 1-125 <FRA>
A:Cross-references: UNIPROT:Q9PSN5; UNIPARC:UPI000013113B
C:Superfamily: Phospholipase A2
F:48,99/Active site: His, Asp #status predicted

Alignment Scores:
Pred. No.: 5.97e-09 Length: 125
Score: 157.00 Matches: 23
Percent Similarity: 75.0% Conservative: 10
Best Local Similarity: 52.3% Mismatches: 11
Query Match: 12.5% Indels: 0
DB: 2 Gaps: 0

US-10-607-806-1-G7328_COPY_7000_7700 (1-701) x S65624 (1-125)
QY 222 CTTTCCAGGTGCTCCAGACACATGACACTGCTATGACAGCCAGAGCTGGACAGC 281
Db 41 LeuAspArgCysGlnThrHisAspAsnCysTyrSerGlyCysGluAlaGluLysLeuProAla 60
QY 282 TGTAAATTTCTGCTGGACAMMCGGTACACCCACCTATTTCATCTGCTGCTGCTGCTG 341
Db 61 CysAsnTyrMetMetSerGlyProTyrTyrAsnThrTyrSerTyrGluCysAsnGluGly 80
QY 342 GCAATCACTGCT 353
Db 81 GluLeuThrCys 84
RESULT 12
PSNJ3K
phospholipase A2 (EC 3.1.1.4) III - monoclod cobra
N:Alternate names: phosphatidylcholine 2-acylhydrolase
C:Species: Naja naja kaouthia, Naja naja siamensis (monocled cobra)
C:Date: 31-Mar-1992 #sequence_revision 14-Feb-1997 #text_change 16-Aug-2004
C:Accession: B00739; A00739
R:Goubert, F.J.; Taljaard, N.
Eur. J. Biochem. 112, 493-499, 1980
A:Title: Purification, some properties and amino-acid sequences of two phospholipases A
A:Reference number: A00739; MUID:81114211; PMID:7460933
A:Accession: B00739
A:Molecule type: protein
A:Residues: 1-70, 'SC', 73-119 <JOU>
A:Cross-references: UNIPARC:UPI0000172763
A:Note: the sequences in Table 9 and Figure 4 disagree with the sequence in Figure 5; th
C:Function:
A:Description: catalyzes hydrolysis of 1,2-diacyl-sn-glycero-3-phosphocholine to 1-acyl-
A:Note: the reaction is strongly enhanced when the phospholipid is condensed into a mic
C:Superfamily: Phospholipase A2
C:Keywords: calcium; carboxylic ester hydrolase; lipid degradation; metalloprotein; pres
F:4,67/Binding site: micellar substrate (Gln, Tyr) #status predicted

F:11-71,26-118,28-44,43-99,50-92,60-85,78-90/Disulfide bonds: #status predicted
F:27,29,31,48/Binding site: calcium (Tyr, Gly, Asp) #status predicted
F:47,93/Active site: His, Asp #status predicted

Alignment Scores:
Pred. No.: 4.14e-08 Length: 119
Score: 149.50 Matches: 28
Percent Similarity: 48.5% Conservative: 5
Best Local Similarity: 41.2% Mismatches: 12
Query Match: 11.9% Indels: 23
DB: 2 Gaps: 2

US-10-607-806-1-G7328_COPY_7000_7700 (1-701) x PSNJ3K (1-119)
QY 222 CTTTCCAGGTGCTCCAGACACATGACACTGCTATGACAGCCAGAGCTGGACAGC 281
Db 40 LeuAspArgCysGlnThrHisAspAsnCysTyrAspGluAlaGluLysLeuSerGly 59
QY 282 TGTAAATTTCTGCTGGACAMMCGGTACACCCACCTATTTCATCTGCTGCTGCTGCTG 341
Db 60 Cys-----TrpProTyrPheLysThrTyrSerTyrGluCysSerGlnGly 74
QY 342 GCAATCACTGCTAGCAGTAGGTTATCCCTCTCTGACCTATGAAATTCATGTTGTTCTC 401
Db 75 ThrLeuThrCysLys----- 79
QY 402 AGTAGCCCGCGGAGAAATAATAGT 425
Db 80 -----GlyGlyAsnAsnAla 84
RESULT 13
PSNJAF
phospholipase A2 (EC 3.1.1.4) precursor [validated] - Chinese cobra
N:Alternate names: phosphatidylcholine 2-acylhydrolase
C:Species: Naja naja atra (Chinese cobra)
C:Date: 29-Jun-1981 #sequence_revision 31-Jan-1997 #text_change 16-Aug-2004
C:Accession: JC2137; IS1017; A00740; S38512
R:Pan, F.M.; Yeh, M.S.; Chang, W.C.; Hung, C.C.; Chiou, S.H.
Biochem. Biophys. Res. Commun. 199, 969-976, 1994
A:Title: Sequence analysis and expression of phospholipase A2 from Taiwan cobra.
A:Reference number: JC2137; MUID:94183285; PMID:7510963
A:Accession: JC2137
A:Molecule type: mRNA
A:Residues: 1-146 <PAN>
A:Cross-references: UNIPROT:P00598; UNIPARC:UPI0000131097; EMBL:X73225; NID:G395191; P1
A:Experimental source: venom gland
R:Pan, F.M.; Chang, W.C.; Chiou, S.H.
Biochem. Mol. Biol. Int. 33, 187-194, 1994
A:Title: cDNA and protein sequences coding for the precursor of phospholipase A2 from T
A:Reference number: IS1017; MUID:94362505; PMID:7521702
A:Accession: IS1017
A>Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: mRNA
A:Residues: 1-21, 'S', 23-75, 'H', 77-90, 'S', 92-146 <PAN2>
A:Cross-references: UNIPARC:UPI00001310B7; EMBL:X77755; NID:G456270; PIDN:CAA54802.1; P1
R:Tsay, I.H.; Wu, S.H.; Lo, T.B.
Toxicol. 19, 141-152, 1981
A:Title: Complete amino acid sequence of a phospholipase A-2 from the venom of Naja naj
A:Reference number: A00740; MUID:81177953; PMID:7222082
A:Accession: A00740
A:Molecule type: protein
A:Residues: 28-110, 'CA', 113-133 'D', 135, 'D', 137-139, 'N', 141-146, 'E' <TSA>
A:Cross-references: UNIPARC:UPI0000172764
R:Chang, L.S.; Kuo, K.W.; Chang, C.C.
Biochim. Biophys. Acta 1202, 216-220, 1993
A:Title: Identification of functional involvement of tryptophan residues in phospholip
A:Reference number: S38512; MUID:94002252; PMID:8399382
A:Accession: S38512
A:Molecule type: protein
A:Residues: 44-54; 84-92 <CHA>
A:Cross-references: UNIPARC:UPI0000172765; UNIPARC:UPI0000172766
A:Note: sequence 1a in Table I has an extra Trp and an incorrect count as compared with
A:Note: Trp-45 is implicated in substrate binding

A: Gene: PLA2
C: Function:
A: Description: catalyzes hydrolysis of 1,2-diacyl-sn-glycero-3-phosphocholine to 1-acyl-
A: Note: the reaction is strongly enhanced when the phospholipid is condensed into a micelle
C: Superfamily: Phospholipase A2
C: Keywords: calcium; carboxylic ester hydrolase; lipid degradation; metalloprotein; vendor
F: 1-27/Domain: signal sequence #status predicted <SIG>
F: 28-146/Product: phospholipase A2 #status experimental <MAT>
F: 31_94/Binding site: micellar substrate (Gln, Tyr) #status predicted
F: 38-98, 53-145, 55-71, 70-126, 77-119, 87-112, 105-117/Disulfide bonds: #status experimental
F: 54, 56, 58, 75/Binding site: calcium (Tyr, Gly, Asp) #status predicted
F: 74/Active site: His #status experimental
F: 120/Active site: Asp #status predicted

Alignment Scores:	
Pred. No.:	1.94e-07
Score:	143.50
Percent Similarity:	63.0%
Best Local Similarity:	52.3%
Query Match:	11.5%
DA:	1
Length:	119
Matches:	24
Conservative:	5
Mismatches:	12
Indels:	1
Gaps:	1

Qy	282	TGTAATTTCTGCTGGACAMCMCGTAGACCCACCACTATTTCATCTCGTGGCTCG	341
Db	60	Cys-----TrpProTyrPheLysThrTyrSerTyrGluCysSerGlnGly	74
Qy	342	GCAATCACCTGACGAGT	359
		:::	
Db	75	ThrLeuThrCysLysAsn	80

RESULT 15

PSOXG

phospholipase A2 (EC 3.1.1.4) taipoxin gamma chain - Australian taipan (tentative sequence)

N;Alternate names: phosphatidylcholine 2-acylhydrolase

C;Species: Oxyuranus scutellatus scutellatus (Australian taipan)

C;Date: 30-Nov-1980 #sequence_revision 30-Nov-1980 #text_change 16-Aug-2004

C;Accession: A00756

R;Fohlman, J.; Lind, P.; Eaker, D.

FEBS Lett. 84, 367-371, 1977

A;Title: Taipoxin, an extremely potent presynaptic snake venom neurotoxin. Elucidation of

A;Reference number: A00756; MUID:78084803; PMID:563806

A;Accession: A00756

A;Molecule type: protein

A;Accession: P100001310D1
A;Cross-references: UNIPROT:P00616; UNIPARC:UPI00001310D1
C;Comment: Taipoxin, a presynaptic neurotoxin of the venom, contains three noncovalently
C;Function:
A;Description: catalyzes hydrolysis of 1,2-diacyl-sn-glycerol-3-phosphocholine to 1-acyl-
A;Note: the reaction is strongly enhanced when the phospholipid is condensed into a micelle
C;Superfamily: phospholipase A2
C;Keywords: calcium; carboxylic ester hydrolase; glycoprotein; lipid degradation; metallo-
E;19-85_23-27_35-131_37-53_52-113_59-106_69-99_92-104/Disulfide bonds: #status predicted
F;36_38_40_57/Binding site: calcium (Tyr, Gly, Asp) #status predicted
F;56_107/Active site: His, Asp #status predicted
F;78/Binding site: carbohydrate (Asn) (covalent) #status experimental

Pred. No.:	2,19e-07	Length:	133
Score:	143.00	Matches:	27
Percent Similarity:	54.0%	Conservative:	7
Best Local Similarity:	42.9%	Mismatches:	19
Query Match:	11.4%	Indels:	10
DB:	1	Gaps:	1
US-10-607-806-1-G7328_COPY_7000_7700 (1-701) x PSOXG (1-133)			
Qy	168	GGTGAGGCCTGACCTATTGCTGTGCACCTTACTCCTATCTCAGTGTCCTCCCACTTTCC	227
Dd	41	GlySerGlyThrProIleAsp-----AspLeuLasp	50

QY 228 AGGTGCTGCCAGACACATGACAACTGCTAYGACCAAGCCAGAGCTGGACAGCTGTAAA 287
 |||||:::||||||| ||||| ::||| ||||| ::||| |||||
 Db 51 ArgCysCysLysThrHisAspGluCysTyrAlaGluAlaGlyLysLeuSerAlaCysLys 70
 QY 288 TTTCCTGCTGGACAMCCGTACACCCACCTATTCTACTGCTGCTGCTGGCGCAATC 347
 ::||| ||| ||||| ||||| ||||| ||||| ||||| ::|||
 Db 71 SerValLeuSerGluProAsnAsnAspThrTyrSerTyrGluCysAsnGluGlyGlnLeu 90
 QY 348 ACCTGTAGC 356
 |||||:::
 Db 91 ThrCysAsn 93

Search completed: February 8, 2006, 11:18:59
 Job time : 25.9191 secs

THIS PAGE BLANK (USPTO)

GenCore version 5.1.7
Copyright (c) 1993 - 2006 Bioceleration Ltd.

OM nucleic - protein search, using frame_plus_n2p model

Run on: February 8, 2006, 10:23:31; Search time 32.3249 Seconds
(without alignments)
3060.024 Million cell updates/sec

Title: US-10-607-806-1-G7328_COPY_7000_7700

Perfect score: 1253

Sequence: 1 gtctgtcactgtgtccag.....tcgagctcaacctcctcgag 701

Scoring table:

BLOSUM62
Xgapop 10.0, Xgapext 0.5
Ygapop 10.0, Ygapext 0.5
Fgapop 6.0, Fgapext 7.0
Delep 6.0, Delext 7.0

Searched: 2166443 seqs, 705528306 residues

Total number of hits satisfying chosen parameters: 4332886

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Command line parameters:

-MODE=frame+n2p model -DEV=xlp
-Q=/abs/ABSSWEB_spool/WARTIN10607806/runat_08022006_092018_27658/app_query.fasta_1
-DB=Uniprot -QFWT=fastan -SUFFIX=n2p.rup -MINMATCH=0.1 -LOOPCL=0 -LOOPEXT=0
-UNITS=bits -START=1 -END=1 -MATRIX=blosum62 -TRANS=human40.cdi -LIST=45
-DOALIGN=200 -THR SCORE=pct -THR MAX=100 -THR MIN=0 -ALIGN=15 -MODE=LOCAL
-OUTFMT=ptc -NORM=ext -HEADSIZE=500 -MINLEN=0 -MAXLEN=2000000000 -HOST=abs02p
-USER=WARTIN10607806 @CCN 1.1.694 @runat_08022006_092018_27658 -NCPU=6 -ICPU=3
-NO MAP -NEG_SCORES=0 -WAIT -DSPBLOCK=100 -LONGLOG -DEV_TIMEOUT=120
-WARN_TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5 -FGAPOP=6 -FGAPEXT=7
-YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database : UniProt 05.80.*

1: uniprot_sprot.*

2: uniprot_trembl.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result	No.	Score	Query Match	Length	DB ID	Description
1	258	20.6	148	1	PA21B_HUMAN	P04054 homo sapien
2	222	17.7	146	1	PA21B_MOUSE	Q920Y2 mus musculus
3	218	17.4	146	1	PA21B_CANFA	P06596 canis famil
4	216	17.2	144	1	PA2_RABIT	Q7m334 oryctolagus
5	216	17.2	146	1	PA21B_RAT	P04055 rattus norv
6	209	16.7	123	1	PA21B_SHEEP	P14419 ovis aries
7	209	16.7	145	1	PA21B_BOVIN	P00593 bos taurus
8	205	16.4	146	1	PA21B_PIG	P00592 sus scrofa
9	202	16.1	124	1	PA22_PIG	P04416 sus scrofa
10	201.5	16.1	132	1	PA21B_HORSE	P00594 equus cabal
11	188	15.0	146	1	PA21B_CAVPO	P43434 cavia porce
12	174	13.9	152	1	PA21_LATSE	O8jfb2 laticauda s
13	170	13.6	152	1	PA21_LATSE	O8jfb2 laticauda s
14	161	12.8	151	1	PA21B_OPHHA	P80966 ophiophagus
15	161	12.8	152	1	PA217_AUSSU	Q9pug7 austrrelaps
16	159	12.7	119	1	PA22_ASPSC	P07037 aspidelaps

17	157	12.5	125	1	PA2E_NOTSC	Q9psn5 notechis sc
18	157	12.5	152	2	Q75848_BUNCA	Q75848 bungarus ca
19	152	12.1	154	1	PA21B_PSETE	Q9w744 pseudonaja
20	149.5	11.9	118	1	PA2H_BUNFA	P29601 bungarus fa
21	148	11.8	154	2	Q4VR15_OXYSC	Q4vr15 oxyuranus s
22	145	11.6	152	1	PA216_AUSSU	Q9pug8 austrrelaps
23	144.5	11.5	146	1	PA21_NAJAT	P00598 naja atra (
24	144.5	11.5	146	1	PA22_NAJKA	P00597 naja kaouth
25	144	11.5	154	1	PA22_PSETE	Q9w743 pseudonaja
26	143	11.4	133	1	PA23_OXYSC	P00616 oxyuranus s
27	143	11.4	152	2	Q4VR16_OXYSC	Q4vr16 oxyuranus s
28	142.5	11.4	146	1	PA22_NAJAT	Q91133 naja atra (
29	142.5	11.4	146	1	PA2A_NAJSP	Q92084 naja sputat
30	142.5	11.4	146	1	PA2B_NAJSP	Q92085 naja sputat
31	142	11.3	152	1	PA23_OPHHA	Q9df56 ophiophagus
32	141.5	11.3	146	1	PA2C_NAJSP	Q92086 naja sputat
33	140.5	11.2	119	1	PA21_HEMHA	P00595 hemachatus
34	140.5	11.2	147	2	Q7TQ04_9SAUR	Q7tq04 bungarus fl
35	139.5	11.1	118	1	PA2X_BUNFA	P14411 bungarus fa
36	139.5	11.1	147	2	Q7TQ05_9SAUR	Q7tq05 bungarus fl
37	139	11.1	117	1	PA2A_PSEPO	P20258 pseudochis
38	138.5	11.1	118	1	PA2B_BUNFA	P00629 bungarus fa
39	138.5	11.1	119	1	PA2_NAJNA	P15445 naja naja (
40	138.5	11.1	120	2	Q65ZF5_NAJNA	Q65zf5 naja naja (
41	138.5	11.1	137	2	Q6SLM2_BUNCE	O6slm2 bungarus ca
42	137.5	11.0	118	1	PA26_BUNFA	P00627 bungarus fa
43	137.5	11.0	118	1	PA2A_BUNFA	P00628 bungarus fa
44	137.5	11.0	126	1	PA22_NAJSG	P60044 naja sagitt
45	137.5	11.0	146	1	PA21_NAJKA	P00596 naja kaouth

ALIGNMENTS

RESULT 1

PA21B_HUMAN	ID	PA21B_HUMAN	STANDARD;	PRT;	148 AA.
AC	P04054;				
DT	01-NOV-1986	(Rel. 03, Created)			
DT	01-FEB-1991	(Rel. 17, Last sequence update)			
DT	13-SEP-2005	(Rel. 48, Last annotation update)			
DE	Phospholipase A2 precursor (EC 3.1.1.4) (Phosphatidylcholine 2-				
DE	acylhydrolase) (Group IB phospholipase A2).				
GN	Names=PLA2G1B; Synonyms=PLA2, PLA2A;				
OS	Homo sapiens (Human)				
OC	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;				
OC	Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini; Hominidae;				
OC	Homo.				
OX	NCBI_TaxID=9606;				
RN	[1]				
RP	NUCLEOTIDE SEQUENCE [GENOMIC DNA / MRNA].				
RC	TISSUE=lung;				
RX	MEDLINE=87132925; PubMed=3028739;				
RA	Seilhamer J.J., Randall T.L., Yamanaka M., Johnson L.K.;				
RT	"Pancreatic phospholipase A2: isolation of the human gene and cDNAs				
RT	from porcine pancreas and human lung.";				
RL	DNA 5:519-527(1986).				
RN	[2]				
RP	NUCLEOTIDE SEQUENCE [GENOMIC DNA].				
RA	Rieder M.J., Livingston R.J., Daniels M.R., Chung M.-W.,				
RA	Miyamoto K.E., Nguyen C.P., Nguyen D.A., Poel C.L., Robertson P.D.,				
RA	Schackwitz W.S., Sherwood J.K., Witrak L.A., Nickerson D.A.;				
RT	"NIH-SNPs, environmental genome project, NIHES ES15478, Department				
RT	of Genome Sciences, Seattle, WA (URL: http://egp.gs.washington.edu).";				
RT	Submitted (Oct-2003) to the EMBL/GenBank/DBJ databases.				
RN	[3]				
RP	NUCLEOTIDE SEQUENCE [LARGE SCALE GENOMIC DNA].				
RL	Human chromosome 12 international sequencing consortium;				
RG	Submitted (JAN-1998) to the EMBL/GenBank/DBJ databases.				
RN	[4]				
RP	PROTEIN SEQUENCE OF 16-22.				
RC	TISSUE=Pancreas;				
RX	MEDLINE=82138816; PubMed=7060561;				
RA	Grataroli R., Dijkman R., Dutilh C.E., van der Ouderaa F.,				

de Haas G.H., Figarella C.;
"Studies on phospholipase A2 and its enzyme from human pancreatic
juice. Catalytic properties and sequence of the N-terminal region.";
Eur. J. Biochem. 122:111-117(1982).
[5]
RN PROTEIN SEQUENCE OF 23-148.
RN TISSUE=Pancreas;
RX MEDLINE=83283533; PubMed=6349696; DOI=10.1016/0167-4838(83)90126-7;
RA Verheij H.M., Westerman J., Sternby B., de Haas G.H.;
RT "The complete primary structure of phospholipase A2 from human
pancreas.";
RL Biochim. Biophys. Acta 747:93-99(1983).
CC -1- FUNCTION: PA2 catalyzes the calcium-dependent hydrolysis of the 2-
acyl groups in 3-sn-phosphoglycerides
CC -1- CATALYTIC ACTIVITY: Phosphatidylcholine + H(2)O = 1-
acylglycerophosphocholine + a carboxylate.
CC -1- COFACTOR: Binds 1 calcium ion per subunit (By similarity).
CC -1- SUBCELLULAR LOCATION: Secreted.
CC -1- SIMILARITY: Belongs to the phospholipase A2 family.
CC This Swiss-Prot entry is copyright. It is produced through a collaboration
between the Swiss Institute of Bioinformatics and the EMBL outstation
the European Bioinformatics Institute. There are no restrictions on its
use as long as its content is in no way modified and this statement is not
removed.
DR EMBL; M21056; AAA60107.1; -; Genomic_DNA.
DR EMBL; M22970; AAA60107.1; JOINED; Genomic_DNA.
DR EMBL; M21054; AAA36450.1; -; mRNA.
DR EMBL; AY438977; AAR05441.1; -; Genomic_DNA.
DR EMBL; AC003982; AAB95635.1; -; Genomic_DNA.
DR PIR; C25793; PSHU.
DR HSSP; P00592; 1HN4.
DR Ensembl; ENSG00000170890; Homo sapiens.
DR HGNC; HGNC:9030; PLA2G1B.
DR MIM; 172410; -.
GO; GO:0004623; P:phospholipase A2 activity; TAS.
GO; GO:0007015; P:actin filament organization; TAS.
GO; GO:0007165; P:signal transduction; TAS.
DR InterPro; IPR001211; PhospholipaseA2.
DR PANTHER; PTHR11716; PhospholipaseA2; 1.
DR Pfam; PF00068; PhospholipaseA2; 1.
DR PRINTS; PR00389; PHPLIPASEA2.
DR ProDom; PD00303; PhospholipaseA2; 1.
DR SMART; SM00085; PA2c; 1.
DR PROSITE; PS00119; PA2_ASP; 1.
DR PROSITE; PS00118; PA2_HIS; 1.
KW Calcium; Direct protein sequencing; Hydrolase; Lipid degradation;
KW Metal-binding; Polymorphism; Signal.
FT SIGNAL 1 15
FT PROPEP 16 22 Activation peptide.
FT CHAIN 23 148 Phospholipase A2.
FT ACT_SITE 70 70 By similarity.
FT ACT_SITE 121 121 Calcium (via carbonyl oxygen) (By
FT METAL 50 50 similarity).
FT METAL 52 52 Calcium (via carbonyl oxygen) (By
FT METAL 54 54 similarity).
FT METAL 54 54 Calcium (By similarity).
FT METAL 71 71
FT DISULFID 33 99
FT DISULFID 49 146
FT DISULFID 51 67
FT DISULFID 66 127
FT DISULFID 73 120
FT DISULFID 83 113
FT DISULFID 106 118
FT VARIANT 16 16 D -> A (in dbSNP:5632).
FT VARIANT 89 89 /FTId=VAR_011911.
FT VARIANT 89 89 N -> K (in dbSNP:5636).
FT VARIANT 89 89 /FTId=VAR_011913.
FT VARIANT 89 89 N -> T (in dbSNP:5635).

FT CONFLICT 144 144 /FTId=VAR_011912.
FT MISSING (in Ref. 5).
SQ SEQUENCE 148 AA; 16360 MW; C8E38B2B64AEE8CB CRC64;
Alignment Scores:
Pred. No.: 4.08e-20 Length: 148
Score: 258.00 Matches: 55
Percent Similarity: 59.8% Conservative: 6
Best Local Similarity: 53.9% Mismatches: 12
Query Match: 20.6% Indels: 29
DB: 1 Gaps: 3
US-10-607-806-1-G7328_COPY_7000_7700 (1-701) x PA21B_HUMAN (1-148)
QY 113 ATTTCTCTCAAGCATCTTGGCGA-----AGT 139
Db 19 ILSerProArgAlaValTrpGlnPheArgLysMetIleLysCysValIleProGlySer 38
QY 140 CATCCCCACCTG-----TTCTGAGAGTGGGCGAGTGGGGCTG 178
Db 39 AsPProPheLeuGluTyrAsnAsnTyrGlyCysTy-CysGlyLeuGly-GlySerGlyTh 58
QY 179 ACTATTGCTCTGCACCTTACTCTATCTAGCTGCTCCCTCCCACTTCCAGGTGCTCCCA 238
Db 58 rProValAsp-----GluLeuAspLysCysCysG 68
QY 239 GACACATGACAACCTGTATGACACGACGAGAGCTGAGAGCTGTAAATTTCTGCTGGA 298
Db 68 nThrHisAspAsnCysTyrAspGlnAlaLysLysLeuAspSerCysLysPheLeuLeuAs 88
QY 299 CMMCCGTACACCCACACCTATTTCATCTGCTCTGCTCGGCAATCATCTGTAGCAG 358
Db 88 pAsnProTyrThrHisThrTyrSerTyrSerCysSerGlySerAlaIleThrCysSerSe 108
QY 359 TAGG 362
Db 108 rLys 109
RESULT 2
PA21B_MOUSE
ID PA21B_MOUSE STANDARD; PRT; 146 AA.
AC Q920V2; Q9D7E2; Q9D884;
DT 16-OCT-2001 (Rel. 40, Created)
DT 16-OCT-2001 (Rel. 40, Last sequence update)
DT 13-SEP-2005 (Rel. 48, Last annotation update)
DE Phospholipase A2 precursor (EC 3.1.1.4) (Phosphatidylcholine 2-
acylhydrolase) (Group IB phospholipase A2) (PLA2-IB).
GN Mus musculus (Mouse).
OS Mus musculus (Mouse).
OC Mammalia; Euthera; Euarchontoglires; Glires; Rodentia; Sciurognathi;
OC Muroidea; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP NUCLEOTIDE SEQUENCE.
RC STRAIN=129/SVJ;
RA Mandal A.K., Zhang Z., Mukherjee A.B.;
RT "Isolation, characterization and chromosomal localization of mouse
sPLA2-IB gene.";
RL Submitted (SEP-1998) to the EMBL/GenBank/DBJ databases.
RN [2]
RP NUCLEOTIDE SEQUENCE.
RC MEDLINE=99167464; PubMed=10066760; DOI=10.1074/jbc.274.11.7043;
RX Cupillard L., Mulharkar R., Gomez N., Kadam S., Valentin E.,
RA Lazdunski M., Lambeau G.;
RT "Both group IB and group IIA secreted phospholipases A2 are natural
ligands of the mouse 180-kDa M-type receptor.";
RL J. Biol. Chem. 274:7043-7051(1999).
RN [3]
RP NUCLEOTIDE SEQUENCE.
RC STRAIN=BALB/C; TISSUE=Lung;
RX MEDLINE=20002676; PubMed=10531350; DOI=10.1074/jbc.274.44.31476;
RA Bingham C.O. III, Fijneman R.J.A., Friend D.S., Goddeau R.P.,

RA Rogers R.A., Austen K.F., Arm J.P.;
RT "Low molecular weight group IIA and group V phospholipase A(2) enzymes
RT have different intracellular locations in mouse bone marrow-derived
RL mast cells.";
RL J. Biol. Chem. 274:31476-31484(1999).
RN [4]
RP NUCLEOTIDE SEQUENCE.
RX MEDLINE=20156393; PubMed=10689188; DOI=10.1016/S0378-1119(00)00006-8;
RA Richmond B.L., Hui D.Y.;
RA "Molecular structure and tissue-specific expression of the mouse
RT pancreatic phospholipase A2 gene.";
RL Gene 244:65-72(2000).
RN [5]
RP NUCLEOTIDE SEQUENCE.
RX STRAIN=C57BL/6J;
RT TISSUE=Pancreas; Small intestine, Stomach, and Tongue;
RC MEDLINE=22354683; PubMed=12468851; DOI=10.1038/nature01266;
RX Okazaki Y., Furuno M., Kasukawa T., Adachi J., Bono H., Kondo S.,
RA Nikaide I., Osato N., Saito R., Suzuki H., Yamanaka I., Kiyosawa H.,
RA Yagi K., Tomaru Y., Hasegawa Y., Nogami A., Schonbach C., Gojobori T.,
RA Baldarelli R., Hill D.P., Bult C., Hume D.A., Quackenbush J.,
RA Schriml L.M., Kanapin A., Matsuda H., Batalov S., Beisel K.W.,
RA Blake J.A., Bradt D., Brusic V., Chochia C., Corbani L.B., Cousins S.,
RA Dalla E., Dragani T.A., Fletcher C.F., Forrest A., Frazer K.S.,
RA Gaasterland T., Gariboldi M., Gissi C., Godzik A., Gough J.,
RA Grimmond S., Gustincich S., Hirokawa N., Jackson I.J., Jarvis E.D.,
RA Kanai A., Kawaji H., Kawasawa Y., Kedzierski R.M., King B.L.,
RA Konagaya A., Kurochkin I.V., Lee Y., Lenhard B., Lyons P.A.,
RA Maglott D.R., Maltais L., Marchionni L., McKenzie L., Miki H.,
RA Nagashima T., Numata K., Okido T., Pavan W.J., Pertea G., Pesole G.,
RA Petrovsky N., Pillai R., Pontius J.U., Qi D., Ramchandran S.,
RA Ravasi T., Reed J.C., Reid D.J., Reid J., Ring B.Z., Ringwald M.,
RA Sandelin A., Schneider C., Semple C.A., Setou M., Shimada K.,
RA Sultana R., Takenaka Y., Taylor M.S., Teasdale R.D., Tomita M.,
RA Virardo R., Wagner L., Wahlestedt C., Wang Y., Watanabe Y., Wells C.,
RA Wilming L.G., Wynshaw-Boris A., Yanagisawa M., Yang I., Yang L.,
RA Yuan Z., Zavalon M., Zhu Y., Zimmer A., Carninci P., Hayatsu N.,
RA Hirozane-Kishikawa T., Konno H., Nakamura M., Sakazume N., Sato K.,
RA Shiraki T., Waki K., Kawai J., Aizawa K., Arakawa T., Fukuda S.,
RA Hara A., Hashizume W., Imotani K., Ishii Y., Itoh M., Kagawa I.,
RA Miyazaki A., Sakai K., Sasaki D., Shibata K., Shinagawa A.,
RA Yasunishi A., Yoshino M., Waterston R., Lander E.S., Rogers J.,
RA Birney E., Hayashizaki Y.,
RT "Analysis of the mouse transcriptome based on functional annotation of
RT 60,770 full-length cDNAs";
RL Nature 420:563-573(2002).
CC -1- FUNCTION: PA2 catalyzes the calcium-dependent hydrolysis of the 2-
CC acyl groups in 3-sn-phosphoglycerides.
CC -1- CATALYTIC ACTIVITY: Phosphatidylcholine + H(2)O = 1-
CC acylglycerophosphocholine + a carboxylate.
CC -1- COFACTOR: Binds 1 calcium ion per subunit (By similarity).
CC -1- SUBCELLULAR LOCATION: Secreted.
CC -1- SIMILARITY: Belongs to the phospholipase A2 family.
CC
CC This Swiss-Prot entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use as long as its content is in no way modified and this statement is not
CC removed.
CC
CC ENBL; AF094611; AAF02298.1; -; Genomic DNA.
CC ENBL; AF097637; AAD19896.1; -; mRNA.
CC ENBL; AF162712; AAD45806.1; -; mRNA.
CC ENBL; AF187852; AAF44297.1; -; mRNA.
CC ENBL; AF094610; AAG27064.1; -; mRNA.
CC ENBL; AK028104; BAC25749.1; -; mRNA.
CC ENBL; AK028134; BAC25763.1; -; mRNA.
DR ENBL; AK008936; BAB25978.1; -; mRNA.
DR ENBL; AK007730; BAB25218.1; -; mRNA.
DR ENBL; AK007797; BAB25263.1; -; mRNA.
DR ENBL; AK008664; BAB25819.1; -; mRNA.
DR ENBL; AK008668; BAB25822.1; -; mRNA.
DR ENBL; AK008841; BAB25922.1; -; mRNA.
DR ENBL; AK008934; BAB25976.1; -; mRNA.
DR ENBL; AK008331; BAB25608.1; -; mRNA.
DR ENBL; AK009314; BAB26212.1; -; mRNA.
DR HSSP; P00593; 1CH4.
DR Ensembl; ENSMUSG00000029522; Mus musculus.
DR MGI; MGI:101842; Pla2g1b.
DR GO; GO:0005615; C:extracellular space; TAS.
DR GO; GO:0004623; F:phospholipase A2 activity; IDA.
DR GO; GO:0005102; F:receptor binding; IDA.
DR GO; GO:0008283; P:cell proliferation; TAS.
DR GO; GO:0009395; P:phospholipid catabolism; TAS.
DR GO; GO:0006950; P:response to stress; TAS.
DR InterPro; IPR01211; PhospholipaseA2.
DR PANTHER; PTHR11716; PhospholipaseA2; 1.
DR Pfam; PF00068; Phospholip A2_1; 1.
DR PRINTS; PR00389; PHPLIPASEA2; 1.
DR ProDom; PD000303; PhospholipaseA2; 1.
DR SMART; SM00085; PA2C; 1.
DR PROSITE; PS00119; PA2 ASP; 1.
DR PROSITE; PS00118; PA2 HIS; 1.
KW Calcium; Hydrolase; Lipid degradation; Metal-binding; Signal.
FT SIGNAL 1 15 By similarity.
FT PROPEP 16 22 Activation peptide (By similarity).
FT CHAIN 23 146 Phospholipase A2.
FT ACT_SITE 70 70 By similarity.
FT ACT_SITE 121 121 By similarity.
FT METAL 50 50 Calcium (via carbonyl oxygen) (By
FT similarity).
FT METAL 52 52 Calcium (via carbonyl oxygen) (By
FT similarity).
FT METAL 54 54 Calcium (via carbonyl oxygen) (By
FT similarity).
FT METAL 71 71 Calcium (By similarity).
FT DISULFID 33 99 By similarity.
FT DISULFID 49 146 By similarity.
FT DISULFID 51 67 By similarity.
FT DISULFID 66 127 By similarity.
FT DISULFID 73 120 By similarity.
FT DISULFID 83 113 By similarity.
FT DISULFID 106 118 By similarity.
FT CONFLICT 52 52 G -> S (in Ref. 6; BAB26212).
FT CONFLICT 58 58 T -> P (in Ref. 6; BAB26212).
FT CONFLICT 78 79 KK -> EN (in Ref. 6; BAB26212).
FT CONFLICT 87 87 I -> R (in Ref. 6; BAB26212).
FT CONFLICT 95 95 Y -> F (in Ref. 6; BAB26212).
FT CONFLICT 102 102 S -> G (in Ref. 6; BAB26212).
FT CONFLICT 122 122 R -> S (in Ref. 6; BAB25608).
SQ SEQUENCE 146 AA; 16290 MW; 59500C6845B7C81 CRC64;
Alignment Scores:
Pred. No.: 5,36e-16 Length: 146
Score: 222.00 Matches: 37
Percent Similarity: 91.5% Conservative: 6
Best Local Similarity: 78.7% Mismatches: 4
Query Match: 17.7% Indels: 0
DB: 1 Gaps: 0
US-10-607-806-1-G7328_COPY_7000_7700 (1-701) x PA21B_MOUSE (1-146)
QY 223 CTTTCCAGTGTGCGACACATGACACTGCTAYGACCGCCAGAGCTGGACGC 281
Db 63 LeuAspArgCysCysGlnThrHisAspHisCysTyrSerGlnAlaLysLysLeuGluSer 82
QY 282 TGTAAATTTCTGCTGGACAMCCGTACACCCACCTATTCATCTACTGCTCTGCTCG 341
Db 83 CysLysPheLeuIleAspAsnProtyrThrAsnThrTyrSerTyrSerCysSerGlySer 102

```
QY 342 GCAATCACCTGACGCTAGG 362
Db 103 GlullethrCysSerAlaLys 109

RESULT 3
PA21B CANFA
ID PA21B CANFA STANDARD; PRT; 146 AA.
AC P06596;
DT 01-JAN-1988 (Rel. 06, Created)
DT 01-JAN-1988 (Rel. 06, Last sequence update)
DT 13-SEP-2005 (Rel. 48, Last annotation update)
DE Phospholipase A2 precursor (EC 3.1.1.4) (Phosphatidylcholine 2-
DE acylhydrolase) (Group IB phospholipase A2).
GN Name=PLA2G1B;
OS Canis familiaris (Dog).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Laurasiatheria; Carnivora; Fissipedia; Canidae;
OC Canis.
OX NCBI_TaxID=9615;
RN [1]
RP NUCLEOTIDE SEQUENCE.
RX MEDLINE=86223862; PubMed=3754861;
RA Ohara O., Tanaki M., Nakamura E., Tsuruta Y., Fujii Y., Shin M.,
RA Teraoka H., Okamoto M.;
RT "Dog and rat pancreatic phospholipases A2: complete amino acid
RT sequences deduced from complementary DNAs.";
RL J. Biochem. 99:733-739(1986).
[2]
RP NUCLEOTIDE SEQUENCE.
RX MEDLINE=90306027; PubMed=2142076;
RA Kerfelec B., Laforge K.S., Vasiloudes P., Puigserver A., Scheele G.A.;
RT "Isolation and sequence of the canine pancreatic phospholipase A2
RT gene.";
RL Eur. J. Biochem. 190:299-304(1990).
[3]
RP NUCLEOTIDE SEQUENCE.
RX MEDLINE=87175472; PubMed=3562437;
RA Kerfelec B., Laforge K.S., Puigserver A., Scheele G.A.;
RT "Primary structures of canine pancreatic lipase and phospholipase A2
RT messenger RNAs.";
RL Pancreas 1:430-437(1986).
CC -!- FUNCTION: PA2 catalyzes the calcium-dependent hydrolysis of the 2-
CC acyl groups in 3-sn-phosphoglycerides.
CC -!- CATALYTIC ACTIVITY: Phosphatidylcholine + H(2)O = 1-
CC acylglycerophosphocholine + a carboxylate.
CC -!- COFACTOR: Binds 1 calcium ion per subunit (By similarity).
CC -!- SUBCELLULAR LOCATION: Secreted.
CC -!- SIMILARITY: Belongs to the phospholipase A2 family.
CC -----
CC This Swiss-Prot entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use as long as its content is in no way modified and this statement is not
CC removed.
CC -----
DR EMBL; D00035; BAA00023.1; -; mRNA.
DR EMBL; M35301; AAA30883.1; -; mRNA.
DR PIR; S11316; PSQG.
DR HSSP; P00592; 1HN4.
DR InterPro; IPR001211; PhospholipaseA2.
DR Ensembl; ENSCAFG0000010263; Canis familiaris.
DR PANTHER; PTHR11716; PhospholipaseA2; 1.
DR Pfam; PF00068; PhospholipA2; 1.
DR PRINTS; PR00389; PHPLIPASEA2.
DR ProDom; PD000303; PhospholipaseA2; 1.
DR SMART; SM00085; PA2C; 1.
DR PROSITE; PS00119; PA2_ASP; 1.
DR PROSITE; PS00118; PA2_HIS; 1.
DR Calcium; Hydrolyase; Lipid degradation; Metal-binding; Signal.
KW SIGNAL
FT PROPEP 16 22
FT CHAIN 23 146 Phospholipase A2.
FT ACT_SITE 70 70 By similarity.

104 ACT_SITE 121 121 By similarity.
105 METAL 50 50 Calcium (via carbonyl oxygen) (By
106 METAL 52 52 similarity).
107 METAL 54 54 Calcium (via carbonyl oxygen) (By
108 METAL 54 54 similarity).
109 METAL 71 71 Calcium (By similarity).
110 DISULFID 33 39 By similarity.
111 DISULFID 49 146 By similarity.
112 DISULFID 51 67 By similarity.
113 DISULFID 66 127 By similarity.
114 DISULFID 73 120 By similarity.
115 DISULFID 83 113 By similarity.
116 DISULFID 106 118 By similarity.
117 SQ SEQUENCE 146 AA; 16236 MW; F6258ED9527F3692 CRC64;

Alignment Scores:
Pred. No.: 1-54e-15 Length: 146
Score: 218.00 Matches: 48
Percent Similarity: 54.9% Conservative: 8
Best Local Similarity: 47.1% Mismatches: 17
Query Match: 17.4% Indels: 29
DB: 1 Gaps: 3

US-10-607-806-1-G7328_COPY_7000_7700 (1-701) x PA21B_CANFA (1-146)
QY 113 ATTTCTCTGAGCATCTTGGCGA-----AGT 139
Db 19 ILeSerProArgAlaValTrpGlnPheArgAsnMetIleLysCysThrIleProGluSer 38
QY 140 CATCCC-----CACCTCTCTCTGAGAGTGGGCGAGTGGGCGCTG 178
Db 39 AspProLeuLysAspTyrAsnAspTyrGlyCysGlyLeuGly-GlySerGlyTh 58
QY 179 ACCTATTGCTCTGCACCTTACTCTATCTCAGCTGTCCCTCCCACTTTTCCAGGTGCTGCCA 238
Db 58 rProValAsp-----GluLeuAspLysCysCysG1 68
QY 239 GACACATGACACTGCTATGACAGCCAGAGAGAGCTGGACAGCTGTAAATTTCTGTGGA 298
Db 68 nThrHisAspHisCysTyrSerGluAlaLysLysLeuAspSerCysLysPheLeuLeuAs 88
QY 299 CAMMCGTACACCCACACCTATTACTCTCTCTGCTCGCTGGCTGGCAATACCTGTAGCAG 358
Db 88 pAsnProTyrThrLysIleTyrSerTyrSerCysSerGlySerGluLeuThrCysSerSe 108
QY 359 TAGG 362
Db 108 rLys 109

RESULT 4
PA2_RABIT
ID PA2_RABIT STANDARD; PRT; 144 AA.
AC Q7M334;
DT 05-JUL-2004 (Rel. 44, Created)
DT 05-JUL-2004 (Rel. 44, Last sequence update)
DT 13-SEP-2005 (Rel. 48, Last annotation update)
DE Phospholipase A2 precursor (EC 3.1.1.4) (Phosphatidylcholine 2-
DE acylhydrolase) (Group IB phospholipase A2).
GN Name=PLA2G1B;
OS Oryctolagus cuniculus (Rabbit).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Euarchontoglires; Glires; Legomorphia; Leporidae;
OC Oryctolagus.
OX NCBI_TaxID=9986;
RN [1]
RP NUCLEOTIDE SEQUENCE.
RX TISSUE=Pancreas;
RX MEDLINE=93249443; PubMed=7916604;
RA Kumar V.B.;
RT "Cloning and expression of rabbit pancreatic phospholipase A2.";
RL Biochem. Biophys. Res. Commun. 192:683-692(1993).
```

CC -!- FUNCTION: PA2 catalyzes the calcium-dependent hydrolysis of the 2-
CC acyl groups in 3-sn-phosphoglycerides (By similarity).
CC -!- CATALYTIC ACTIVITY: Phosphatidylcholine + H(2)O = 1-
CC acylglycerophosphocholine + a carboxylate.
CC -!- COFACTOR: Binds 1 calcium ion per subunit (By similarity).
CC -!- SUBCELLULAR LOCATION: Secreted (By similarity).
CC -!- SIMILARITY: Belongs to the phospholipase A2 family.
CC -----
CC This Swiss-Prot entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use as long as its content is in no way modified and this statement is not
CC removed.
CC -----
CC PIR; JN0480; JN0480.
CC InterPro; IPR001211; PhospholipaseA2.
CC PANTHER; PTHR11716; PhospholipaseA2; 1.
CC Pfam; PF00068; Phospholip A2_1; 1.
CC PRINTS; PR00389; PHPLIPASEA2.
CC ProDom; PD000303; PhospholipaseA2; 1.
CC PROSITE; PS00119; PA2_ASP; 1.
CC PROSITE; PS00118; PA2_HIS; 1.
CC Calcium; Hydrolase; Lipid degradation; Metal-binding; Signal.
CC SIGNAL 1 15 By similarity.
CC PROPEP 16 22 Activation peptide.
CC CHAIN 23 144 Phospholipase A2.
CC ACT_SITE 70 70 By similarity.
CC ACT_SITE 121 121 By similarity.
CC METAL 50 50 Calcium (via carbonyl oxygen) (By
CC similarity).
CC METAL 52 52 Calcium (via carbonyl oxygen) (By
CC similarity).
CC METAL 54 54 Calcium (via carbonyl oxygen) (By
CC similarity).
CC METAL 71 71 Calcium (By similarity).
CC DISULFID 33 99 By similarity.
CC DISULFID 49 144 By similarity.
CC DISULFID 51 67 By similarity.
CC DISULFID 66 127 By similarity.
CC DISULFID 73 120 By similarity.
CC DISULFID 83 113 By similarity.
CC DISULFID 106 118 By similarity.
CC SQ SEQUENCE 144 AA; 15733 MW; DF404C1F28AE932A CRC64;
Alignment Scores:
Pred. No.: 2,6e-15 Length: 144
Score: 216.00 Matches: 45
Percent Similarity: 55.9% Conservative: 12
Best Local Similarity: 44.1% Mismatches: 16
Query Match: 17.2% Indels: 29
DB: 3 Gaps: 3
US-10-607-806-1-G7328_COPY_7000_7700 (1-701) x PA2_RABIT (1-144)
QY 113 ATTTCTCTGAAGCATCTTGGCGA-----AGT 139
Db 19 ValSerProThrAlaLeuTrpGlnPheArgGlyMetIleGlnCysThrIleProGlySer 38
QY 140 CATCCACACTG-----TTCTGAGATGGGCGAGGTGAGGGCTG 178
Db 39 SerProTyrLeuGluPheAsnGlyTyrGlyCysTyrCysGlyLeuGly-GlySerGlyTh 58
QY 179 ACCTATTGCTCTGCACCTTACTCTATCTAGCTGTCTCCCTCCCACTTCCAGGTGCTGCCA 238
Db 58 rProValAsp-----GluLeuAspArgCysCysG 68
QY 239 GACACATGACAACTGCTAYGACCAGGCCAAGAGCTGGACAGCTGTAAATTTCTGCTGGA 298
Db 68 nThrHisAspGlnCysTyrThrGlnAlaLysLysLeuSerSerCysSerPheLeuValAs 88
QY 299 CAMMCCGTACACCCACACCTATTATCATCTGCTCTGCTGGCTGGCAATCACCTGTAGCAG 358
Db 88 pAsnProTyrThrAsnSerTyrSerTyrSerCysSerGlyThrThrValThrCysSerSe 108

QY 359 TAGG 362
Db 108 rLys 109
RESULT 5
PA21B RAT STANDARD; PRT; 146 AA.
ID PA21B RAT
AC P04055;
DT 01-NOV-1986 (Rel. 03, Created)
DT 01-NOV-1986 (Rel. 03, Last sequence update)
DE 13-SEP-2005 (Rel. 48, Last annotation update)
DE Phospholipase A2 precursor (EC 3.1.1.4) (Phosphatidylcholine 2-
DE acylhydrolase) (Group IB phospholipase A2).
GN Name=Pla2g1b;
OS Rattus norvegicus (Rat).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Euarchontoglires; Glires; Rodentia; Sciurognathi;
OC Muridae; Muridae; Muridae; Muridae; Muridae;
OX NCBI_TaxID=101116;
RN [1]
RP NUCLEOTIDE SEQUENCE.
RX MEDLINE=86223862; PubMed=3754861;
RA Ohara O., Tamaki M., Nakamura E., Tsuruta Y., Fujii Y., Shin M.,
RA Teraoka H., Okamoto M.;
RT "Dog and rat pancreatic phospholipases A2: complete amino acid
RT sequences deduced from complementary DNAs.";
RL J. Biochem. 99:733-739 (1986).
RN [2]
RP NUCLEOTIDE SEQUENCE.
RX MEDLINE=89076874; PubMed=2909239; DOI=10.1016/0167-4781(89)90141-3;
RA Sakata T., Nakamura E., Tsuruta Y., Tamaki M., Teraoka H., Tojo H.,
RA Ono T., Okamoto M.;
RT "Presence of pancreatic-type phospholipase A2 mRNA in rat gastric
RT mucosa and lung.";
RL Biochim. Biophys. Acta 1007:124-126 (1989).
RN [3]
RP PROTEIN SEQUENCE OF 23-54.
RX MEDLINE=85054750; PubMed=6501264;
RA Ono T., Tojo H., Inoue K., Kagamiyama H., Yamano T., Okamoto M.;
RT "Rat pancreatic phospholipase A2: purification, characterization, and
RT N-terminal amino acid sequence.";
RL J. Biochem. 96:785-792 (1984).
CC -!- FUNCTION: PA2 catalyzes the calcium-dependent hydrolysis of the 2-
CC acyl groups in 3-sn-phosphoglycerides.
CC -!- CATALYTIC ACTIVITY: Phosphatidylcholine + H(2)O = 1-
CC acylglycerophosphocholine + a carboxylate.
CC -!- COFACTOR: Binds 1 calcium ion per subunit (By similarity).
CC -!- SUBCELLULAR LOCATION: Secreted.
CC -!- SIMILARITY: Belongs to the phospholipase A2 family.
CC -----
CC This Swiss-Prot entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use as long as its content is in no way modified and this statement is not
CC removed.
CC -----
CC EMBL; D00036; BAA00024.1; -; mRNA.
CC PIR; A92008; PSRT.
CC HSSP; P00593; 4BP2.
CC Ensembl; ENSRNOG0000001153; Rattus norvegicus.
CC RGD; 61949; Pla2g1b.
CC InterPro; IPR001211; PhospholipaseA2.
CC PANTHER; PTHR11716; PhospholipaseA2; 1.
CC Pfam; PF00068; Phospholip A2_1; 1.
CC PRINTS; PR00389; PHPLIPASEA2.
CC ProDom; PD000303; PhospholipaseA2; 1.
CC SMART; SM00085; PA2C; 1.
CC PROSITE; PS00119; PA2_ASP; 1.
CC PROSITE; PS00118; PA2_HIS; 1.
CC Calcium; Direct protein sequencing; Hydrolase; Lipid degradation;
CC Metal-binding; Signal.
CC SIGNAL 1 15

```

FT PROPEP      16 22      Activation peptide.
FT CHAIN       23 146      Phospholipase A2.
FT ACT_SITE   70 70
FT ACT_SITE  121 121
FT METAL      50 50
FT METAL      52 52      Calcium (via carbonyl oxygen) (By
FT METAL      54 54      similarity).
FT METAL      71 71      Calcium (via carbonyl oxygen) (By
FT DISULFID   33 99      similarity).
FT DISULFID   49 146      By similarity.
FT DISULFID   51 67      By similarity.
FT DISULFID   66 127      By similarity.
FT DISULFID   73 120      By similarity.
FT DISULFID   83 113      By similarity.
FT DISULFID  106 118      By similarity.
SQ SEQUENCE  146 AA; 16424 MW; 7EC4F7A91B913D0 CRC64;

Alignment Scores:
Pred. No.:      2,61e-15      Length:      146
Score:          216.00      Matches:      35
Percent Similarity: 89.4%      Conservative: 7
Best Local Similarity: 74.5%      Mismatches: 5
Query Match:    17.2%      Indels:      0
DB:             1      Gaps:      0

US-10-607-806-1-G7328_COPY_7000_7700 (1-701) x PA21B_RAT (1-146)

QY 222 CTTTCCAGGTCGCCAGACATGACAACTGCTAYGACCGCCAGAGCTGGACAGC 281
Db 63 LeuAspArgCysGlnThrHisAspHisCysTyrAsnGlnAlaLysLeuGluSer 82

QY 282 TGTAAATTTCTGCTGGACAMMCCGTACACACCTATTTCATCTGCTCTGGCTCG 341
Db 83 CysLysPheLeuIleAspAsnProTyrThrAsnThrTyrSerTyrLysCysSerGlyAsn 102

QY 342 GCATCACCCTTACAGTAGG 362
Db 103 ValIleThrCysSerAspLys 109

RESULT 6
PA21B SHEEP
ID PA21B_SHEEP STANDARD; PRT; 123 AA.
AC P14419;
DT 01-JAN-1990 (Rel. 13, Created)
DT 01-JAN-1990 (Rel. 13, Last sequence update)
DT 13-SEP-2005 (Rel. 48, Last annotation update)
DE Phospholipase A2 (EC 3.1.1.4) (Phosphatidylcholine 2-acylhydrolase)
DE (Group IB phospholipase A2).
GN Name=PLA2G1B;
OS Ovis aries (Sheep).
OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Laurasiatheria; Cetartiodactyla; Ruminantia;
OC Pecora; Bovidae; Caprinae; Ovis.
OX NCBI_TaxID=9940;
RN [1]
RP PROTEIN SEQUENCE.
RC TISSUE=Pancreas;
RA Verheij H.M., Raykova D., Dijkman R., Lenting H.B.M., de Haas G.H.;
RT "The primary structure and some kinetic properties of phospholipase A2
RL from sheep pancreas.";
RL Recl. Trav. Chim. Pays-Bas 107:203-207(1988).
CC -!- FUNCTION: PA2 catalyzes the calcium-dependent hydrolysis of the 2-
CC acyl groups in 3-sn-phosphoglycerides.
CC -!- CATALYTIC ACTIVITY: Phosphatidylcholine + H(2)O = 1-
CC acylglycerophosphocholine + a carboxylate.
CC -!- COFACTOR: Binds 1 calcium ion per subunit (By similarity).
CC -!- SUBCELLULAR LOCATION: Secreted.
CC -!- SIMILARITY: Belongs to the phospholipase A2 family.
CC -----
CC This Swiss-Prot entry is copyright. It is produced through a collaboration

```

```

CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use as long as its content is in no way modified and this statement is not
CC removed.
CC -----
CC HSSP; P00593; 4BP2.
CC InterPro; IPR001211; PhospholipaseA2.
CC PANTHER; PTHR11716; PhospholipaseA2; 1.
CC Pfam; PF00068; PhospholipA2_1; 1.
CC PRINTS; PR00389; PHPLIPASEA2.
CC ProDom; PD000303; PhospholipaseA2; 1.
CC SMART; SM00085; PA2c; 1.
CC PROSITE; PS00119; PA2_ASP; 1.
CC PROSITE; PS00118; PA2_HIS; 1.
KW Calcium; Direct protein sequencing; Hydrolase; Lipid degradation;
KW Metal-binding.
FT ACT_SITE 48 48      By similarity.
FT ACT_SITE 99 99      By similarity.
FT METAL    28 28      Calcium (via carbonyl oxygen) (By
FT METAL    30 30      similarity).
FT METAL    32 32      Calcium (via carbonyl oxygen) (By
FT METAL    49 49      similarity).
FT DISULFID 11 77      Calcium (By similarity).
FT DISULFID 27 123      By similarity.
FT DISULFID 29 45      By similarity.
FT DISULFID 44 105      By similarity.
FT DISULFID 51 98      By similarity.
FT DISULFID 61 91      By similarity.
FT DISULFID 84 96      By similarity.
SQ SEQUENCE 123 AA; 13817 MW; F69878C15DF6C4F9 CRC64;

Alignment Scores:
Pred. No.:      1.6e-14      Length:      123
Score:          209.00      Matches:      36
Percent Similarity: 84.8%      Conservative: 3
Best Local Similarity: 78.3%      Mismatches: 7
Query Match:    16.7%      Indels:      0
DB:             1      Gaps:      0

US-10-607-806-1-G7328_COPY_7000_7700 (1-701) x PA21B_SHEEP (1-123)

QY 222 CTTTCCAGGTCGCCAGACATGACAACTGCTAYGACCGCCAGAGCTGGACAGC 281
Db 41 LeuAspArgCysGlnThrHisAspAsnCysTyrLysGlnAlaLysLysLeuAspSer 60

QY 282 TGTAAATTTCTGCTGGACAMMCCGTACACACCTATTTCATCTGCTCTGGCTCG 341
Db 61 CysLysValLeuValAspAsnProTyrThrAsnSerTyrSerTyrSerCysSerAsnLys 80

QY 342 GCATCACCCTTACAGTAGG 359
Db 81 GlnIleThrCysSerSer 86

RESULT 7
PA21B BOVIN
ID PA21B_BOVIN STANDARD; PRT; 145 AA.
AC P00593;
DT 21-JUL-1986 (Rel. 01, Created)
DT 16-OCT-2001 (Rel. 40, Last sequence update)
DT 13-SEP-2005 (Rel. 48, Last annotation update)
DE Phospholipase A2 precursor (EC 3.1.1.4) (Phosphatidylcholine 2-
DE acylhydrolase) (Group IB phospholipase A2).
GN Name=PLA2G1B;
OS Bos taurus (Bovine).
OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Laurasiatheria; Cetartiodactyla; Ruminantia;
OC Pecora; Bovidae; Bovinae; Bos.
OX NCBI_TaxID=9913;
RN [1]
RP NUCLEOTIDE SEQUENCE.

```

RX MEDLINE=87174835; PubMed=3562249;
RA Tanaka T., Kimura S., Ota Y.;
RT "Sequence of a cDNA coding for bovine pancreatic phospholipase A2.";
RL Nucleic Acids Res. 15:3178-3178(1987).
RN [2]
RP PROTEIN SEQUENCE OF 16-22.
RC TISSUE=Pancreas;
RA Dutliff C.B., van Doren P.J., Verheul F.E.A.M., de Haas G.H.;
RT "Isolation and properties of phospholipase A2 from ox and sheep
pancreas.";
RL Eur. J. Biochem. 53:91-97(1975).
RN [3]
RP PROTEIN SEQUENCE OF 23-145.
RA MEDLINE=78084411; PubMed=620674;
RC Fleer E.M., Verheul F.E.A.M., de Haas G.H.;
RT "The primary structure of bovine pancreatic phospholipase A2.";
RL Eur. J. Biochem. 82:261-269(1978).
RN [4]
RP X-RAY CRYSTALLOGRAPHY (2.4 ANGSTROMS), AND DISULFIDE BONDS.
RA MEDLINE=79050564; PubMed=712836;
RC Dijkstra B.W., Drenth J., Kalk K.H., Vandermaelen P.J.;
RT "Three-dimensional structure and disulfide bond connections in bovine
pancreatic phospholipase A2.";
RL J. Mol. Biol. 124:53-60(1978).
RN [5]
RP X-RAY CRYSTALLOGRAPHY (1.7 ANGSTROMS), ACTIVE SITE, AND CATALYTIC
MECHANISM.
RA MEDLINE=81123082; PubMed=7464926;
RC Dijkstra B.W., Drenth J., Kalk K.H.;
RT "Active site and catalytic mechanism of phospholipase A2.";
RL Nature 289:604-606(1981).
RN [6]
RP X-RAY CRYSTALLOGRAPHY (1.7 ANGSTROMS).
RA MEDLINE=81267392; PubMed=7265241;
RC Dijkstra B.W., Kalk K.H., Hol W.G.J., Drenth J.;
RT "Structure of bovine pancreatic phospholipase A2 at 1.7A resolution.";
RL J. Mol. Biol. 147:97-123(1981).
RN [7]
RP X-RAY CRYSTALLOGRAPHY (3.0 ANGSTROMS) OF PROENZYME.
RA Dijkstra B.W., van Nes G.J.H., Kalk K.H., Brandenburg N.P.,
Hol W.G.J., Drenth J.;
RT "The structure of bovine pancreatic phospholipase A2 at 3.0-A
resolution.";
RL Acta Crystallogr. B 38:793-799(1982).
RN [8]
RP X-RAY CRYSTALLOGRAPHY (1.72 ANGSTROMS).
RA MEDLINE=9728082; PubMed=9115986; DOI=10.1021/bi961576x;
RC Sekar K., Yu B.Z., Rogers J., Lutton J., Liu X., Chen X., Tsai M.-D.,
Jain M.K., Sundaralingam M.;
RT "Phospholipase A2 engineering. Structural and functional roles of the
highly conserved active site residue aspartate-99.";
RL Biochemistry 36:3104-3114(1997).
RN [9]
RP X-RAY CRYSTALLOGRAPHY (1.5 ANGSTROMS).
RA MEDLINE=98039151; PubMed=9369492; DOI=10.1021/bi971370b;
RC Sekar K., Eswaramoorthy S., Jain M.K., Sundaralingam M.;
RT "Crystal structure of the complex of bovine pancreatic phospholipase
A2 with the inhibitor 1-hexadecyl-3-(trifluoroethyl)-sn-glycero-2-
phosphomethanol.";
RL Biochemistry 36:14186-14191(1997).
RN [10]
RP X-RAY CRYSTALLOGRAPHY (1.95 ANGSTROMS).
RA MEDLINE=99190898; PubMed=10093353; DOI=10.1107/S0907444998013699;
RC Sekar K., Blasas R., Li Y., Tsai M., Sundaralingam M.;
RT "Structures of the catalytic site mutants D99A and H48Q and the
calcium-loop mutant D49E of phospholipase A2.";
RL Acta Crystallogr. D 55:443-447(1999).
RN [11]
RP STRUCTURE BY NMR.
RA MEDLINE=99175086; PubMed=10074343; DOI=10.1021/bi9822123;
RC Yuan C., Byeon I.-J.L., Li Y., Tsai M.-D.;
RT "Structural analysis of phospholipase A2 from functional perspective.
1. Functionally relevant solution structure and roles of the hydrogen-
bonding network.";
RL Biochemistry 38:2909-2918(1999).
RN [12]
RP STRUCTURE BY NMR OF MUTANTS.
RA MEDLINE=99175087; PubMed=10074344; DOI=10.1021/bi9822123;
RC Yuan C., Byeon I.-J.L., Li Y., Tsai M.-D.;
RT "Structural analysis of phospholipase A2 from functional perspective.
2. Characterization of a molten globule-like state induced by site-
specific mutagenesis.";
RL Biochemistry 38:2919-2929(1999).
RN [13]
RP FUNCTION: PA2 catalyzes the calcium-dependent hydrolysis of the 2-
acyl groups in 3-sn-phosphoglycerides.
CC -1- CATALYTIC ACTIVITY: Phosphatidylcholine + H(2)O = 1-
acylglycerophosphocholine + a carboxylate.
CC -1- COFACTOR: Binds 1 calcium ion per subunit.
CC -1- SUBCELLULAR LOCATION: Secreted.
CC -1- SIMILARITY: Belongs to the phospholipase A2 family.
CC
CC This Swiss-Prot entry is copyright. It is produced through a collaboration
between the Swiss Institute of Bioinformatics and the EMBL Outstation -
the European Bioinformatics Institute. There are no restrictions on its
use as long as its content is in no way modified and this statement is not
removed.
CC
CC EMBL: Y00120; CAA68303.1; -; mRNA.
DR PIR; A27508; PSBOA.
DR PDB; 1BP2; X-ray; @=23-145.
DR PDB; 1BPQ; X-ray; @=23-145.
DR PDB; 1BVN; NMR; A=23-145.
DR PDB; 1C74; X-ray; A=23-145.
DR PDB; 1CBH; X-ray; @=23-145.
DR PDB; 1CEH; X-ray; @=23-145.
DR PDB; 1G41; X-ray; A=23-145.
DR PDB; 1G44; X-ray; A=23-141.
DR PDB; 1IIR; X-ray; @=23-141.
DR PDB; 1KVN; X-ray; @=23-145.
DR PDB; 1KVV; X-ray; @=23-145.
DR PDB; 1KVS; X-ray; @=23-145.
DR PDB; 1MKT; X-ray; @=23-145.
DR PDB; 1MKU; X-ray; @=23-145.
DR PDB; 1O2E; X-ray; A=23-145.
DR PDB; 1O3W; X-ray; A=23-145.
DR PDB; 1UNE; X-ray; @=23-145.
DR PDB; 1VQV; X-ray; A=23-145.
DR PDB; 1VL9; X-ray; A=23-144.
DR PDB; 2BP2; X-ray; @=17-145.
DR PDB; 2BPP; X-ray; @=23-145.
DR PDB; 3BP2; X-ray; @=24-145.
DR PDB; 4BP2; X-ray; @=16-145.
DR InterPro; IPR001211; PhospholipaseA2.
DR PANTHER; PTHR11716; PhospholipaseA2; 1.
DR Pfam; PF00068; Phospholip A2 1; 1.
DR PRINTS; PR00389; PHPLIPASEA2.
DR ProDom; PD000303; PhospholipaseA2; 1.
DR SMART; SM00085; PA2C; 1.
DR PROSITE; PS00119; PA2 ASP; 1.
DR PROSITE; PS00118; PA2 HIS; 1.
DR 3D-structure; Calcium; Direct protein sequencing; Hydrolase;
KW Lipid degradation; Metal-binding; Pyrolidone carboxylic acid; Signal.
FT SIGNAL 1 15
FT PROPEP 16 22 Removed by trypsin.
FT CHAIN 23 145 Phospholipase A2.
FT ACT SITE 70 70
FT ACT SITE 121 121
FT METAL 50 50 Calcium (via carbonyl oxygen).
FT METAL 52 52 Calcium (via carbonyl oxygen).
FT METAL 54 54 Calcium (via carbonyl oxygen).
FT METAL 71 71 Calcium.
FT MOD RES 16 16 Pyrolidone carboxylic acid.
FT DISULFID 33 99
FT DISULFID 49 145

```
FT DISULFID 51 67
FT DISULFID 66 127
FT DISULFID 73 120
FT DISULFID 83 113
FT DISULFID 106 118
FT DISULFID 144 144
FT CONFLICT 24 34
FT HELIX 36 37
FT TURN 40 43
FT TURN 44 44
FT STRAND 47 47
FT STRAND 48 50
FT STRAND 51 51
FT STRAND 62 79
FT HELIX 81 85
FT TURN 86 87
FT STRAND 90 92
FT STRAND 97 100
FT TURN 101 102
FT STRAND 103 106
FT TURN 108 109
FT HELIX 112 129
FT TURN 130 130
FT HELIX 135 137
FT STRAND 138 138
FT TURN 139 139
FT HELIX 142 145
SQ SEQUENCE 145 AA; 16002 MW; 17A1E04B0C22F668 CRC64;

Alignment Scores:
Pred. No.: 1.65e-14 Length: 145
Score: 209.00 Matches: 36
Percent Similarity: 84.8% Conservative: 3
Best Local Similarity: 78.3% Mismatches: 7
Query Match: 16.7% Indels: 0
DB: 1 Gaps: 0

US-10-607-806-1-G7328_COPY_7000_7700 (1-701) x PA21B_BOVIN (1-145)
Qy 222 CTTTCAGGCTGCTGCCAGACATGACATCTGTAYGACCAAGCCAGAGCTGGACAGC 281
Db 63 LeuAspArgCysGlnThrHisAspAsnCysTyrLysGlnAlaLysLysLeuAspSer 82
Qy 282 TGTAAATTCCTGCGGACAMWCCGTACACACACCTATTTCATCTCTGCTCTGGCTCG 341
Db 83 CysLysValLeuAlaAspAsnProTyrThrAsnAsnTyrSerTyrSerCysSerAsn 102
Qy 342 GCAATCACCTGTAGCAGT 359
Db 103 GluIleThrCysSerSer 108

RESULT 8
PA21B_PIG
ID PA21B_PIG STANDARD; PRT; 146 AA.
AC P00592;
DT 21-JUL-1986 (Rel. 01, Created)
DT 01-JAN-1988 (Rel. 06, Last sequence update)
DT 13-SEP-2005 (Rel. 48, Last annotation update)
DE Phospholipase A2, major isoenzyme precursor (EC 3.1.1.4)
DE (Phosphatidylcholine 2-acylhydrolase) (Group 1B phospholipase A2).
GN Name=PLA2G1B;
OS Sus scrofa (pig)
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Laurasiatheria; Cetartiodactyla; Suina; Suidae;
OC Sus.
OC NCBI_TaxID=9823;
RN [1]
RP NUCLEOTIDE SEQUENCE.
RC TISSUE=Pancreas;
RA de Geus P., van den Bergh C.J., Kuipers O., Verheij H.M.,
RA Hoekstra W.P.M., de Haas G.H.;
RT "Expression of porcine pancreatic phospholipase A2. Generation of
active enzyme by sequence-specific cleavage of a hybrid protein from
Escherichia coli.";
RL Nucleic Acids Res. 15:3743-3759(1987).
[2]
RN RP NUCLEOTIDE SEQUENCE.
RC TISSUE=Pancreas;
RA MEDLINE=87132925; PubMed=3028739;
RA Seilhamer J.J., Randall T.L., Yamanaka M., Johnson L.K.;
RT "Pancreatic phospholipase A2: isolation of the human gene and cDNAs
from porcine pancreas and human lung.";
RL DNA 5:519-527(1986).
[3]
RN RP PROTEIN SEQUENCE OF 16-146.
MEDLINE=71014235; PubMed=5528841;
RA de Haas G.H., Slotboom A.J., Bonsen P.P.M., van Deenen L.L.M.,
RA Maroux S., Puigserver A., Desnuelle P.;
RT "Studies on phospholipase A and its zymogen from porcine pancreas. I.
The complete amino acid sequence.";
RL Biochim. Biophys. Acta 221:31-53(1970).
[4]
RN RP SEQUENCE REVISION.
MEDLINE=7722092; PubMed=884127; DOI=10.1016/0005-2795(77)90076-9;
RA Puijk W.C., Verheij H.M., de Haas G.H.;
RT "The primary structure of phospholipase A2 from porcine pancreas. A
reinvestigation.";
RL Biochim. Biophys. Acta 492:254-259(1977).
[5]
RN RP DISULFIDE BONDS.
MEDLINE=71014236; PubMed=4919729;
RA de Haas G.H., Slotboom A.J., Bonsen P.P.M., Nieuwenhuizen W.,
RA van Deenen L.L.M., Maroux S., Dlouha V., Desnuelle P.;
RT "Studies on phospholipase A and its zymogen from porcine pancreas. II.
The assignment of the position of the six disulfide bridges.";
RL Biochim. Biophys. Acta 221:54-61(1970).
[6]
RN RP ACYLATION OF LYS-78.
TISSUE=Pancreas;
MEDLINE=89255488; PubMed=2498336;
RA Tomaselli A.G., Hui J., Fisher J., Zuercher-Neely H., Reardon H.M.,
RA Oriaku E., Kezdy F.J., Heinrichson R.L.;
RT "Dimerization and activation of porcine pancreatic phospholipase A2
via substrate level acylation of lysine 56.";
RL J. Biol. Chem. 264:10041-10047(1989).
[7]
RN RP X-RAY CRYSTALLOGRAPHY (2.6 ANGSTROMS), AND ACTIVE AND BINDING SITES.
MEDLINE=83268704; PubMed=6876174;
RA Dijkstra B.W., Renetseder R., Kalk K.H., Hol W.G.J., Drenth J.;
RT "Structure of porcine pancreatic phospholipase A2 at 2.6-A resolution
and comparison with bovine phospholipase A2.";
RL J. Mol. Biol. 168:163-179(1983).
[8]
RN RP X-RAY CRYSTALLOGRAPHY (2.1 ANGSTROMS).
MEDLINE=91073405; PubMed=2254938;
RA Thunnissen M.M.G.M., Kalk K.H., Drenth J., Dijkstra B.W.;
RT "Structure of an engineered porcine phospholipase A2 with enhanced
activity at 2.1-A resolution. Comparison with the wild-type porcine
and Crotaus atrox phospholipase A2.";
RL J. Mol. Biol. 216:425-439(1990).
[9]
RN RP X-RAY CRYSTALLOGRAPHY (2.4 ANGSTROMS).
MEDLINE=91015391; PubMed=2215698; DOI=10.1038/347689a0;
RA Thunnissen M.M.G.M., Ab E., Kalk K.H., Drenth J., Dijkstra B.W.,
RA Kuipers O.P., Dijkman R., de Haas G.H., Verheij H.M.;
RT "X-ray structure of phospholipase A2 complexed with a substrate-
derived inhibitor.";
RL Nature 347:689-691(1990).
[10]
RN RP STRUCTURE BY NMR.
MEDLINE=91175768; PubMed=2007145;
RA Dekker N., Peters A.R., Slotboom A.J., Boelens R., Kaptein R.,
RA de Haas G.H.;
RT "Porcine pancreatic phospholipase A2: sequence-specific 1H and 15N NMR
assignments and secondary structure.";
```

Biochemistry 30:3135-3147(1991).
 [11]
 RN STRUCTURE BY NMR.
 RP MEDLINE=95393214; PubMed=7664098;
 RA van den Berg B., Tessari M., Boelens R., Dijkman R., de Haas G.H.,
 RA Kaptein R., Verheij H.M.;
 RT "NMR structures of phospholipase A2 reveal conformational changes
 during interfacial activation.";
 RL Nat. Struct. Biol. 2:402-406(1995).
 RN [12]
 RN STRUCTURE BY NMR.
 RP MEDLINE=96016184; PubMed=7556053;
 RA van den Berg B., Tessari M., de Haas G.H., Verheij H.M., Boelens R.,
 RA Kaptein R.;
 RT "Solution structure of porcine pancreatic phospholipase A2.";
 RL EMBO J. 14:4123-4131(1995).
 RN [13]
 RN STRUCTURE BY NMR.
 RP MEDLINE=95218441; PubMed=7703697;
 RA van den Berg B., Tessari M., Boelens R., Dijkman R., Kaptein R.,
 RA de Haas G.H., Verheij H.M.;
 RT "Solution structure of porcine pancreatic phospholipase A2 complexed
 with micelles and a competitive inhibitor.";
 RL J. Biomol. NMR 5:110-121(1995).
 CC -1- FUNCTION: PA2 catalyzes the calcium-dependent hydrolysis of the 2-
 CC acyl groups in 3-sn-phosphoglycerides.
 CC -1- CATALYTIC ACTIVITY: Phosphatidylcholine + H(2)O = 1-
 CC acylglycerophosphocholine + a carboxylate.
 CC -1- COFACTOR: Binds 1 calcium ion per subunit.
 CC -1- SUBUNIT: Monomer or homodimer.
 CC -1- SUBCELLULAR LOCATION: Secreted.
 CC -1- PTM: Acylation causes dimerization.
 CC -1- MISCELLANEOUS: Loss of activity upon alkylation of His-70 with p-
 CC bromo phenacyl bromide; Ca(2+) and Ba(2+) protect against
 CC inactivation.
 CC -1- SIMILARITY: Belongs to the phospholipase A2 family.

Alignment Scores:
 Pred. No.: 4,73e-14 Length: 146
 Score: 205.00 Matches: 44
 Percent Similarity: 53.9% Conservative: 11
 Best Local Similarity: 43.1% Mismatches: 18
 Query Match: 16.4% Indels: 29
 DB: 1 Gaps: 3

US-10-607-806-1-G7328_COPY_7000_7700 (1-701) x PA21B_PIG (1-146)

QY 113 ATTTCTCTGAAGCATCTTGGCGA-----AGT 139
 ||||| ||| |||||
 Db 19 IleSerSerArgAlaLeuTrpGlnPheArgSerMetIleLysCysAlaIleProGlySer 38
 ||||| |||||
 QY 140 CATCCCCACCTG-----TTCCTGAGAGTGGCAGGTGAGGGCTG 178
 ||||| |||||
 Db 39 HisProLeuMetAspPheAsnAsnTyrGlyCysTyrCysGlyLeuGly-GlySerGlyTh 58
 ||||| |||||
 QY 179 ACCTATTGTCTCTGCACCTTACTCTCTATCTCAGCTGTCCCTCCACCTTTCCAGGTGCTGCCA 238
 ||||| |||||
 Db 58 rProValasp-----GluLeuAspArgCysCysGly 68
 ||||| |||||
 QY 239 GACACATGACAACTGCTATGACCGCCAGAGCTGGACAGCTGTAATTTCTGCTGGA 298
 ||||| |||||
 Db 68 uThrHisAspAsnCysTyrArgAspAlaLysAsnLeuAspSerCysLysPheLeuValAs 88
 ||||| |||||
 QY 299 CAMMCCGTACACCCACACCTATTCTACTCTGCTGCTGCTGGCAATCACCTGTAGCAG 358
 ||||| |||||
 Db 88 pAsnProTyrThrGluSerTyrSerTyrSerCysSerAsnThrGluIleThrCysAsnse 108
 ||||| |||||

QY = 359 TAGG 362
 |||||
 Db 108 rlys 109

RESULT 9
 PA22_PIG
 ID PA22_PIG STANDARD; PRT; 124 AA.
 AC P04416;


```

QY 137 -----AGTCATCCCCAC-----CTGTTCTCG 157
Db 32 LysCysAlaalleProGlySerArgProTyrSerGluTyrAsnAsnTyrGlyCysPheCys 51
QY 158 AGAGTGGCAGGTGAGGCTGACCTATTGCTCTGCTGCACTTACTCTATCTCAGCTGTCCCT 217
Db 52 GlyLeuGly-GlySerGlyThrProValasp----- 61
QY 218 CCACACTTCCAGGCTGCTGCAGACACATGACACTGCTGCTAYCACCAGCCAGAGCTGGA 277
Db 62 -GluLeuAspArgCysCysGluIleHisAspAlaCysTyrThrGlnAlaLysHisLeuG1 81
QY 278 CAGCTGTAATTTCTGCTGCAGCMMCCGTACCCACACCTATTCATCTGCTGCTG 337
Db 81 userCysLysSerValleAspAsnProTyrThrAsnSerTyrSerPheSerCysSerG1 101
QY 338 CTCGGCAATCACCCTGTAGCAGTAGG 362
Db 101 YThrAsnIleIleCysSerLys 109

RESULT 12
PA2H_LATSE STANDARD; PRT; 152 AA.
AC Q8JFB2;
DT 28-FEB-2003 (Rel. 41, Created)
DT 28-FEB-2003 (Rel. 41, Last sequence update)
DT 13-SEP-2005 (Rel. 48, Last annotation update)
DE Phospholipase A2 GLi6-1 precursor (EC 3.1.1.4) (Phosphatidylcholine 2-
DE acylhydrolase) (pkp5).
OS Laticauda semifasciata (Broad-banded blue sea snake) (Erabu sea
OS snake).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Lepidosauria; Squamata; Scleroglossa; Serpentes; Colubroidea;
OC Elapidae; Laticaudinae; Laticauda.
OX NCBI_TaxID=8631;
RN [1]
RP NUCLEOTIDE SEQUENCE.
RC TISSUE=Pancreas;
RX MEDLINE=22114984; PubMed=12119117; DOI=10.1016/S0378-1119(02)00682-0;
RA Fujimi T.J., Kariya Y., Tsuchiya T., Tamiya T.;
RT "Nucleotide sequence of phospholipase A2 gene expressed in snake
RT pancreas reveals the molecular evolution of toxic phospholipase A2
RT Genes.";
RL Gene 292:225-231(2002).
CC -!- FUNCTION: PA2 catalyzes the calcium-dependent hydrolysis of the 2-
CC acyl groups in 3-sn-phosphoglycerides (By similarity).
CC -!- CATALYTIC ACTIVITY: Phosphatidylcholine + H(2)O = 1-
CC acylglycerophosphocholine + a carboxylate.
CC -!- COFACTOR: Binds 1 calcium ion per subunit (By similarity).
CC -!- SUBCELLULAR LOCATION: Secreted (By similarity).
CC -!- SIMILARITY: Belongs to the phospholipase A2 family. Group I
CC subfamily.
-----
CC This Swiss-Prot entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use as long as its content is in no way modified and this statement is not
CC removed.
-----
CC EMBL; AB078348; BAC03247.1; -; mRNA.
CC DR EMBL; AB078346; BAC03245.1; -; Genomic_DNA.
CC DR HSSP; P00593; 4BP2.
CC DR InterPro; IPR001211; PhospholipaseA2.
CC DR PANTHER; PTHR11716; PhospholipaseA2; 1.
CC DR Pfam; PF00068; PHPLIPASEA2.
CC DR PRINTS; PR00389; PHPLIPASEA2.
CC DR ProDom; PD000303; PhospholipaseA2; 1.
CC DR SMART; SM00085; PA2c; 1.
CC DR PROSITE; PS00119; PA2_ASP; 1.
CC DR PROSITE; PS00118; PA2_HIS; 1.
CC DR Calcium; Hydrolase; Lipid degradation; Metal-binding;
CC KW Multigene family; Signal.
CC SIGNAL 1 21

```

```

FT PROPEP 22 27 By similarity.
FT CHAIN 28 152 Phospholipase A2 GLi6-1.
FT ACT_SITE 75 75 By similarity.
FT ACT_SITE 126 126 By similarity.
FT METAL 55 55 Calcium (via carbonyl oxygen) (By
FT similarity).
FT METAL 57 57 Calcium (via carbonyl oxygen) (By
FT similarity).
FT METAL 59 59 Calcium (via carbonyl oxygen) (By
FT similarity).
FT METAL 76 76 Calcium (By similarity).
FT DISULFID 38 104 By similarity.
FT DISULFID 54 151 By similarity.
FT DISULFID 56 72 By similarity.
FT DISULFID 71 132 By similarity.
FT DISULFID 78 125 By similarity.
FT DISULFID 88 118 By similarity.
FT DISULFID 111 123 By similarity.
SQ SEQUENCE 152 AA; 16430 MW; B86D54DC4B7732D CRC64;

Alignment Scores:
Pred. No.: 1.68e-10 Length: 152
Score: 174.00 Matches: 29
Percent Similarity: 75.0% Conservative: 4
Best Local Similarity: 65.9% Mismatches: 11
Query Match: 13.9% Indels: 0
DB: 1 Gaps: 0

US-10-607-806-1-G7328_COPY_7000_7700 (1-701) x PA2H_LATSE (1-152)
QY 222 CTTTCCAGGTGCTGCCAGACACATGACACTGCTAYCACCAGCCAGAGCTGGACAGC 281
Db 68 LeuAspArgCysCysGlnThrHisAspAsnCysTyrSerGlnAlaLysLysHisProAla 87
QY 282 TGTAAATTTCTGCTGCACMMCCGTACCCACACCTATTCATCTGCTGCTCTGGCTCG 341
Db 88 CysLysSerProLeuAspSerProTyrIleLysIleTyrSerTyrThrCysSerGlyGly 107
QY 342 GCAATCACCTGT 353
Db 108 SerLeuThrCys 111

RESULT 13
PA2H_LATSE STANDARD; PRT; 152 AA.
AC Q8JFB2;
DT 28-FEB-2003 (Rel. 41, Created)
DT 28-FEB-2003 (Rel. 41, Last sequence update)
DT 13-SEP-2005 (Rel. 48, Last annotation update)
DE Phospholipase A2 pkP2 precursor (EC 3.1.1.4) (Phosphatidylcholine 2-
DE acylhydrolase).
OS Laticauda semifasciata (Broad-banded blue sea snake) (Erabu sea
OS snake).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Lepidosauria; Squamata; Scleroglossa; Serpentes; Colubroidea;
OC Elapidae; Laticaudinae; Laticauda.
OX NCBI_TaxID=8631;
RN [1]
RP NUCLEOTIDE SEQUENCE.
RC TISSUE=Pancreas;
RX MEDLINE=22114984; PubMed=12119117; DOI=10.1016/S0378-1119(02)00682-0;
RA Fujimi T.J., Kariya Y., Tsuchiya T., Tamiya T.;
RT "Nucleotide sequence of phospholipase A2 gene expressed in snake
RT pancreas reveals the molecular evolution of toxic phospholipase A2
RT Genes.";
RL Gene 292:225-231(2002).
CC -!- FUNCTION: PA2 catalyzes the calcium-dependent hydrolysis of the 2-
CC acyl groups in 3-sn-phosphoglycerides (By similarity).
CC -!- CATALYTIC ACTIVITY: Phosphatidylcholine + H(2)O = 1-
CC acylglycerophosphocholine + a carboxylate.
CC -!- COFACTOR: Binds 1 calcium ion per subunit (By similarity).
CC -!- SUBCELLULAR LOCATION: Secreted (By similarity).
CC -!- SIMILARITY: Belongs to the phospholipase A2 family. Group I
CC subfamily.
-----
CC This Swiss-Prot entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use as long as its content is in no way modified and this statement is not
CC removed.
-----
CC EMBL; AB078348; BAC03247.1; -; mRNA.
CC DR EMBL; AB078346; BAC03245.1; -; Genomic_DNA.
CC DR HSSP; P00593; 4BP2.
CC DR InterPro; IPR001211; PhospholipaseA2.
CC DR PANTHER; PTHR11716; PhospholipaseA2; 1.
CC DR Pfam; PF00068; PHPLIPASEA2.
CC DR PRINTS; PR00389; PHPLIPASEA2.
CC DR ProDom; PD000303; PhospholipaseA2; 1.
CC DR SMART; SM00085; PA2c; 1.
CC DR PROSITE; PS00119; PA2_ASP; 1.
CC DR PROSITE; PS00118; PA2_HIS; 1.
CC DR Calcium; Hydrolase; Lipid degradation; Metal-binding;
CC KW Multigene family; Signal.
CC SIGNAL 1 21

```


GenCore version 5.1.7
Copyright (c) 1993 - 2006 Bioceleration Ltd.

OM nucleic - protein search, using frame_plus_n2p model

Run on: February 8, 2006, 09:43:51 ; Search time 14.7309 Seconds
(without alignments)
2613.961 Million cell updates/sec

Title: US-10-607-806-1_COPY_7200_9200
Perfect score: 3524
Sequence: 1 ctatctcagctgtccctccc.....ckctacataaatatctatta 2001

Scoring table: BLOSUM62
Xgapop 10.0 , Xgapext 0.5
Ygapop 10.0 , Ygapext 0.5
Fgapop 6.0 , Fgapext 7.0
Delop 6.0 , Delext 7.0

Searched: 283416 seqs, 96216763 residues

Total number of hits satisfying chosen parameters: 565832

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Command line parameters:
-MODEL=frame+ n2p.model -DEV=xlp
-O=/abss/ABSSWEB_spool/US10607806/runat_08022006_091821_26876/app_query.fasta_1
-DB=PIR -QFMT=fastan -SUFFIX=n2p.rpr -MINMATCH=0.1 -LOOPCL=0 -LOOPEXT=0
-UNITS=bits -START=1 -END=1 -MATRIX=blosum62 -TRANS=human40.cdi -LIST=45
-DOALIGN=200 -THR SCORE=pct -THR MAX=100 -THR MIN=0 -ALIGN=15 -MODE=LOCAL
-OUTFMT=ptc -NORM=ext -HAPSIZE=500 -MINLEN=0 -MAXLEN=2000000000 -HOST=abss02p
-USER=US10607806 @CGN 1 1 92 @runat_08022006_091821_26876 -NCPU=6 -ICPU=3
-NO WMAP -NEG SCORES=0 -WAIT -DSPBLOCK=100 -LONGLOG -DEV TIMEOUT=120
-WARN TIMEOUT=30 -THRSADS=1 -XGAPOP=10 -XGAPEXT=0.5 -FGAPOP=6 -FGAPEXT=7
-YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database : PIR 80.*
1: pir1.*
2: pir2.*
3: pir3.*
4: pir4.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	416	11.8	673	4 F40201	artifact-warning s
C 2	391.5	11.1	673	4 F40201	artifact-warning s
C 3	384.5	10.9	613	4 C40201	artifact-warning s
4	366	10.4	613	4 C40201	artifact-warning s
5	332.5	9.4	627	4 A40201	artifact-warning s
C 6	322	9.1	627	4 A40201	artifact-warning s
C 7	267	7.6	597	4 E40201	artifact-warning s
8	264	7.5	196	2 I38022	hypothetical prote
9	259.5	7.4	597	4 E40201	artifact-warning s
10	251	7.1	148	1 PSHU	phospholipase A2 (
C 11	248.5	7.1	301	4 B40201	artifact-warning s
12	239.5	6.8	301	4 B40201	artifact-warning s
13	236.5	6.7	579	4 D40201	artifact-warning s
C 14	232.5	6.6	100	2 A46010	X-linked retinopat

15	227	6.4	418	2 S41044	chromosomal protei
C 16	223.5	6.3	579	4 D40201	artifact-warning s
17	217	6.2	146	1 PSDG	phospholipase A2 (
18	216	6.1	146	1 PSRT	phospholipase A2 (
C 19	214.5	6.1	499	2 S65657	alpha-1C-adrenergi
20	209	5.9	145	1 PSBOA	phospholipase A2 (
C 21	208	5.9	144	2 JN0480	phospholipase A2 (
C 22	206.5	5.9	418	2 S41044	chromosomal protei
C 23	206	5.8	196	2 I38022	hypothetical prote
24	202	5.7	124	1 PSFGA2	phospholipase A2 (
25	201	5.7	146	1 PSFGA	phospholipase A2 (
26	191.5	5.4	132	1 PSHOA	phospholipase A2 (
C 27	177	5.0	146	1 S34049	gene NF2 protein -
C 28	172	4.9	46	2 I54375	thromboxane A-2 re
C 29	162	4.6	39	2 I54374	thromboxane A-2 re
C 30	160	4.5	369	2 A53959	phospholipase A2 (
C 31	157	4.5	125	2 S65624	integrin beta-1 ch
C 32	156	4.4	53	2 A42442	phospholipase A2 (
C 33	155.5	4.4	119	2 S00227	probable thromboxa
C 34	155.5	4.4	407	2 T02670	thromboxane A-2 re
C 35	154.5	4.4	79	2 A56194	phospholipase A2 (
C 36	149.5	4.2	119	1 PSNJ3K	serine/threonine-s
C 37	147	4.2	841	1 I78885	phospholipase A2 (
38	144.5	4.1	146	1 PSNJAF	alpha-1C-adrenergi
39	144.5	4.1	499	2 S65657	phospholipase A2 (
C 40	143.5	4.1	119	1 PSNJ2K	B-cell growth fact
C 41	143.5	4.1	124	2 A47582	X-linked retinopat
C 42	142.5	4.0	100	2 A46010	phospholipase A2 (
43	142	4.0	133	1 PSOXG	phospholipase A2 (
44	140.5	4.0	119	1 PSRIA	phospholipase A2 (
45	138.5	3.9	118	1 PSKFT3	phospholipase A2 (

ALIGNMENTS

RESULT 1

F40201
artifact-warning sequence (translated ALU class F) - human
C:Species: Homo sapiens (man)
C:Date: 31-Mar-1992 #sequence_revision 11-Aug-1995 #text_change 19-May-2000
C:Accession: F40201
R:Claverie, J.M.
personal communication, 1992
A:Reference number: A40201
A:Accession: F40201
A:Molecule type: DNA
A:Residues: 1-673 <CLA>
A:Cross-references: UNIPARC:UPI000017CECD
R:Claverie, J.M.
Genomics 12, 838-841, 1992
A:Title: Identifying coding exons by similarity search: Alu-derived and other potential
A:Reference number: A40200; MUID:92241891; PMID:1572661
A:Contents: annotation
C:Comment: This "warning" entry is a conceptual translation in all 6 reading frames of
in-frame stop codons are shown as 'X'.
C:Comment: Any significant similarity of a predicted protein sequence to a portion of r

Alignment Scores:			
Pred. No.:	2,89e-35	Length:	673
Score:	416.00	Matches:	131
Percent Similarity:	41.8%	Conservative:	17
Best Local Similarity:	37.0%	Mismatches:	85
Query Match:	11.8%	Indels:	122
DB:	4	Gaps:	9

US-10-607-806-1_COPY_7200_9200 (1-2001) x F40201 (1-673)

Qy	332	AAGAAACTGAGGCTCAGAGGCTGACGATTTGCCCAAGATCACACAGCTTGTAACTGG	391
Db	328	LysLysLysAsnProGlnLys-----Gln*****PheCysPheTrp	342
Qy	392	TGACAGTTTGGGTTTTTTTTTTTGTGTGTTTGA-----GAGACAGGGTCTTGCTCTGTC	444

Db 343 GlyPhePhePhePhePheCysPheValPheValPheGluIleGlyProCysSerAsp 362
QY 445 ACCAGGCGATCAGCAGTGTGTGCAACCATAGGTCACTGAGCTCAACCTCTCGAGCTC 504
Db 363 ThrProAlaGlyValGlnTrp-GlnValLeuAlaHisCysSerLeuAsnLeuLeuGlySe 382
QY 505 AAGGAGCTGCTGACCTCAGCCTCCCAAGTACGTGGGACTACGAGCGGTGCACCAACCGC 564
Db 382 rSerAspSerProAlaSerValSerArgValAlaGlyIleThrGlyMetGlyArgHisSe 402
QY 565 CTGGCTAATTAATAAATTTTTGTAGAGACTGGTCTTACTAGTGGCCAGCTGTGT 624
Db 402 rTrpLeuIleTrpVal--PheLeuIleGluThrGlnPheHisValAspGlnAlaGly 421
QY 625 CTAAACTCTGGCTTCAAGCAATCTCTACCTTGGCATCCCAAGTGTGGGATTACA 684
Db 422 LeuLysLeuLeuThrSerSerAspLeuProSer-TrpSerProLysValLeuGly**G1 441
QY 685 GGGGTGAGCCACC-----ATGTCGGCTACTTATTCTTTA 720
Db 441 nAla***AlaThrThrProSer*****PheValPheGlyGlyPhePhePhePh 461
QY 721 CATTCCA----- 727
Db 461 ePheAlaLeuPheLeuPheLeuArg**AlaLeuAlaLeuThrProArgLeuGluCysSe 481
QY 728 ----TCTTTCCAATAGATGTAAAGATCCACA-----GAACAGGATTACTGCTTAT 774
Db 481 rGlyLysPheTrpLeuThrAlaAlaSerThrSerTrpValGlnAlaIleLeuLeuProLe 501
QY 775 T-----TTCTTCCTTTC 786
Db 501 uSerProVal**LeuGlyLeuGlnAlaTrpAlaAlaIleProGly***PheMetTyzPh 521
QY 787 TTTTGTGACAGAGTCTCACT-----TCACTACCTCAA-- 820
Db 521 e*****ArgHisSerPheThrMetLeuIleArgLeuValSerAsnSer***ProGlnVa 541
QY 821 -----CCTCCG----- 826
Db 541 lIleCysProProGlyLeuProLysCysTrpAspAspArgArgGluProProHisProAl 561
QY 826 ----- 826
Db 561 a*****LeuPheLeuGlyValPhePhePhePheLeuLeuCysPheCysPhe** 581
QY 827 -----TTCAGCTCACTGCA 840
Db 581 *AspArgProLeuLeu***HisProGlyTrpSerAlaValAlaSerPheGlySerLeuG1 601
QY 841 ACCTCTGCTCCCGGCTCAAGYGATTCTCTGCTAAGCTCTCCTGAGTAGCTGGAATTA 900
Db 601 nProGlnProProGlyPheLysArgPheSerCysLeuCysLeuProCysSerTrpAspTr 621
QY 901 CAAGCTGCACCAACCATCTGTGCTAATTTTTGTATTTTTTAGCAGAGATGGGGTTTTAC 960
Db 621 rArgHisGlyProProPheLeuAlaAsn-LeuCysIlePheAsnArgAspThrValSerP 641
QY 961 CATGTTGCCAGGCTGTCTCAAACTCCTGACCTCAAGTGATCTGCTGCTGCTCAGTCTCC 1020
Db 641 roCys***SerGlyTrpSerGlnThrProAspLeuLys***SerAlaLeuLeu-ValSer 660
QY 1021 CAAGTGTCTGGAATTATAGCGGTGAGTCAAC 1050
Db 661 GlnSerAlaGlyMetThrGlyValSerHis 670

RESULT 2
F40201
artifact-warning sequence (translated ALU class F) - human
C;Species: Homo sapiens (man)
C;Date: 31-Mar-1992 #sequence_revision 11-Aug-1995 #text_change 19-May-2000
C;Accession: F40201

R;Claverie, J.M.
Personal communication, 1992
A;Reference number: A40201
A;Accession: F40201
A;Molecule type: DNA
A;Residues: 1-673 <CLA>
A;Cross-references: UNIPARC:UPI000017CECD
R;Claverie, J.M.
Genomics 12, 838-841, 1992
A;Title: Identifying coding exons by similarity search: Alu-derived and other potential
A;Reference number: A40200; MUID:92241891; PMID:1572661
C;Contents: annotation
C;Comment: This "warning" entry is a conceptual translation in all 6 reading frames of o
in-frame stop codons are shown as 'X'.
C;Comment: Any significant similarity of a predicted protein sequence to a portion of th

Alignment Scores:
Pred. No.: 1,15e-32 Length: 673
Score: 391.50 Matches: 122
Percent Similarity: 57.9% Conservative: 21
Best Local Similarity: 49.4% Mismatches: 65
Query Match: 11.1% Indels: 41
Caps: 5

US-10-607-806-1_COPY_7200_9200 (1-2001) x F40201 (1-673)
QY 1046 CTCAGCGCTATAATTCAGACACTTTGGGAGACTGAGCAGGCAGATCACTTGAGTCAGG 987
Db 6 LeuThrProValIleProAlaLeuTrpGluThr-ArgArgAlaAspHisLeuArgSerG1 25
QY 986 AGTTTGACGACGCTGGGCAACATGTAACCCCATCTCTGCTAAATAATACAAAAAT 927
Db 25 yVal***AspGlnProAspGlnHisGlyGluThrValSerLeuLeuLysIleHis-LysL 45
QY 926 TAGCAAGCATGTGTGTGACGCTTTGTAATTCAGCTACTCTCAGAGGCTTAGGCGAGGA 867
Db 45 euAlaArgAsnGlyProCysLeu**SerGlnLeuHisGlyArgGlnArgGlnGluA 65
QY 866 ATCCTTTGAACCCGGGAGGAGAGTTGCAGTAG------CTGAACG 825
Db 65 snArgLeuAsnProGlyGly**GlyCysSerGluProLysLeuAlaThrAlaLeuGlnP 85
QY 824 GAGCTTG-AGGTGATGAGTCAGACTCTGCTCAAAAAGAAAGAGAAATAGGCAGT 766
Db 85 roGlyCysGln-----SerLysGlyLeuSerGlnLysGlnLysGlnSerLysLys- 102
QY 765 AATCCTCTGCTGTGGATCTTACATCTTATTGGAAAGATGGAATGTAAAGAAATAAGTA- 707
Db 103 -----LysLysLysThrP 107
QY 706 -----GCCGCACATGTGGTGCCTACCCCTGTATAATCCCGACACTTTCG 667
Db 107 roLysAsnLys*****AlaGlyCysGlyGlySerArgLeuSerSerGlnHisPheG 127
QY 666 GGATCCCAAGGTAGAGGATTCCTTGAAGCCAGGAGTTTAAGACAAGCTGCGCAACGTA 607
Db 127 lYargPro-GlyGlyGlnIleThr***GlyGlnGluPheGluThrSerLeuIleAsnMet 146
QY 606 GTAAGACCCAGTCTCTACAAAAAATTTTTTAATTAGCCAGCGCTGTGTGTGCACGCTC 547
Db 147 ValLysLeuCysLeuTyz**Lys--TyrIleAsn***ProGlyMetAlaAlaHisAlaC 166
QY 546 GTAGTCCCAGCTACTTTGGAGGCTGAGGTACAGATCCCTGAGCTCAGGAGGTTGAGG 487
Db 166 ysAsnProSerTyzThrGlyAspArgGlyArgGileAla***ThrGlnGluValGluA 186
QY 486 CTGCAGTGACCTATGTGTCACCACTGTGCTCATGCTGAGTGCAGACAGACAGACCTGT 427
Db 186 laAlaValSerGlnAsn-LeuProLeuHisSerSerArgGlyValArgAlaArgAlaTy 205
QY 426 CTCTA-----AACCAACAACAAAAAACCCAACTGTCTCACTTACAGCTGTGTG 374

Db 206 LeuLysAsnLysAsnLysAlaLysLysLysLysLysLysProProLysThrLys***** 225

QY 373 ATCTTGGC 365

Db 226 ***LeuGly 228

RESULT 3

C40201

artifact-warning sequence (translated ALU class C) - human

C:Species: Homo sapiens (man)

C:Date: 31-Mar-1992 #sequence_revision 11-Aug-1995 #text_change 19-May-2000

C:Accession: C40201

R:Claverie, J.M.

personal communication, 1992

A:Reference number: A40201

A:Accession: C40201

A:Molecule type: DNA

A:Residues: 1-613 <CLA>

A:Cross-references: UNIPARC:UPI000017CECA

R:Claverie, J.M.

Genomics 12, 838-841, 1992

A:Title: Identifying coding exons by similarity search: Alu-derived and other potential

A:Reference number: A40200; MUID:92241891; PMID:1572661

A:Contents: annotation

C:Comment: This "warning" entry is a conceptual translation in all 6 reading frames of c

in-frame stop codons are shown as 'X'.

C:Comment: Any significant similarity of a predicted protein sequence to a portion of th

Alignment Scores:

Pred. No.:	6,29e-32	Length:	613
Score:	384.50	Matches:	117
Percent Similarity:	52.0%	Conservative:	24
Best Local Similarity:	43.2%	Mismatches:	79
Query Match:	10.9%	Indels:	51
DB:	4	Gaps:	4

US-10-607-806-1_COPY_7200_9200 (1-2001) x C40201 (1-613)

QY 1081 AAGAAATAGACGTAATCGCGAGGACAGTACTCAGCGCTAATATTCAGCACTTT 1022

Db 97 LysLysLysLys*****AlaGlyCysSerGlySerArgLeu***SerGlnHisPhe 116

QY 1021 GGGAGACTGAGCGAGCAGATCATTGAGGTGAGGATTTGAGACCGAGCTGGGCAACAT 962

Db 117 GlyArgGlnArgTTPValAspHis--GluAlaArgSerSerArgProAlaTrpProThr 136

QY 961 GGTAAACCCCATCTCTCTTAAATATACAAAATATGCCAAGCATGTGTGACGCTT 902

Db 136 rpGlnAsnProIleSerThrLysAsnThrLysAsn***LeuGlyMetVal-ArgAlaPro 155

QY 901 GTAATTCAGCTACTCAGGAGCTTAGCGAGGAGATCCTTGAACCGGGAGGACAGG 842

Db 156 ValValProArgThrGlnHisAlaGluAlaGluTrpHisLysProGlyArgTrpSer 175

QY 841 TTGCAGTGAGCTGAA-----CGGAGTTGAGGTGATGATGAGTGCAGACTCTG 797

Db 176 LeuGln***AlaGluIleSerProLeuHisSerSerLeu-GlyAspArgValArgLeuCy 195

QY 796 TCTCAAAAAGAGGAAGAAATAGGAGTAATCCCTGTTCTGTGGATCTTACATTCTA 737

Db 195 sLeuLysLysLysLysLysLys----- 202

QY 736 TTGGAAGATGGAATGTAAGAATAAGTAGCCGCACATGCTGGCTCACCCCTGTAATCC 677

Db 203 -----Lys*****ArgGlyAlaValAlaHisAlaCyAsnPr 216

QY 676 CAGCACTTTGGATGCCAAGGTAGGAGGATTTGTTGAAGCAGGAGTTTAAAGACAAGCT 617

Db 216 oSerThrLeuGlyLysGly-GlyTrpIle-MetArgProGlyValArgAspGlnPro 235

QY 616 GGCAACGTAAGACCCAGTCTCTACAAAAAATTTTTTAATAGCCAGGCGTGTG 557

Db 616 ----- 557

Db 236 GlyGlnHisGlyLysThrProPheLeu-LeuLysIleGlnLysIleSerTrpAlaTrpCy 255

QY 556 GTGCACGCTCGTAGTCCCGAGCTACTTGGAGGCTGAGGTACAGATCCCTTGAGCTCAG 497

Db 255 s-GlyArgLeu***SerHisValArgArgArgLeuArgGlnGlnGlnGlyIleAsnProG 275

QY 496 GAGTTGAGGCTGAGTCCAGTCCCTATGTTGACCACTGCTCATGCTGGGTGACAGAGC 437

Db 275 lyGlyGlyAlaCysSerGluProArgSerArgHisCysThrProAlaTrpValThrGlu* 295

QY 436 AAGACCTGCTCTTAAACAAACAAACAAAAA----- 405

Db 295 **AspSerValSerLysLysLysLysLys*****PhePhePhePhePhePhePheP 315

QY 404 -----ACCCAAACTGTCCACCACTTACAAAGCTGTGTGATCTTG 368

Db 315 heGluThrGluSerHisSerValThrGlnAlaGlyValGlnTrpArg-----AspLeuG 333

QY 367 GGCAAAATAGTCCAGCCT 351

Db 333 lySerLeuGlnAlaPro 338

RESULT 4

C40201

artifact-warning sequence (translated ALU class C) - human

C:Species: Homo sapiens (man)

C:Date: 31-Mar-1992 #sequence_revision 11-Aug-1995 #text_change 19-May-2000

C:Accession: C40201

R:Claverie, J.M.

personal communication, 1992

A:Reference number: A40201

A:Accession: C40201

A:Molecule type: DNA

A:Residues: 1-613 <CLA>

A:Cross-references: UNIPARC:UPI000017CECA

R:Claverie, J.M.

Genomics 12, 838-841, 1992

A:Title: Identifying coding exons by similarity search: Alu-derived and other potentia

A:Reference number: A40200; MUID:92241891; PMID:1572661

A:Contents: annotation

C:Comment: This "warning" entry is a conceptual translation in all 6 reading frames of

in-frame stop codons are shown as 'X'.

C:Comment: Any significant similarity of a predicted protein sequence to a portion of i

Alignment Scores:	5,77e-30	Length:	613
Pred. No.:	366.00	Matches:	134
Score:	40.8%	Conservative:	17
Percent Similarity:	36.2%	Mismatches:	129
Best Local Similarity:	10.4%	Indels:	90
Query Match:	4	Gaps:	6
DB:			

US-10-607-806-1_COPY_7200_9200 (1-2001) x C40201 (1-613)

QY 199 CTAGTAGCGCGGGGAAATAATAGTAACAACAGCCATGATTTAGTGTAAATTTCTTG 258

Db 226 IleMetArgProGlyValArgAspGlnProGlyGlnHisGlyLysThrProPheLeuLeu 245

QY 259 GTTCTGGCAGTGTCTCTTTTAAATCTCTCAGACAACACTATGGGATAGGTACAAATTATCC 318

Db 246 LysIleGlnLysIleSerTrpAlaTrpCysGlyArgLeu***SerHisValArgArg 265

QY 319 TCATTAAACAGATAAGAAACTGAGGCTCAGAGGCTGAGTATTTGCCCAAGATCACAC 378

Db 266 Leu-----ArgGlnGluAsnGlyIleAsnProGlyGlyAlaCysSerGluProArg 283

QY 379 AGCTTGTAAGTGGTGACAGTTTGGTT----- 405

Db 284 SerArgHisCysThrProAlaTrpValThrGlu****AspSerValSerLysLysLys 303

QY 406 -----TTTTTTTGTGTTTGTAGACAGAGGCTTCTGCTGTGTCACC 447

Db 406 ----- 447

Db 304 LysIys*****PhePhePhePhePheGluThrGluSerHisSerValThr 323
 QY 448 CAGCATGAGCACAGTGTGCAACATAGTCTCAGTGCAGCTCAACCTCTCGACTCAAG 507
 Db 324 GlnAlaGlyValGlnTrpArgAspLeuGlySerLeuGlnAlaProProGlyPheMet 343
 QY 508 GATCTGCTGACCTCAGCTCCAGTACGTGGGACTAGCGCTGACACACCGCTG 567
 Db 344 ProPheSerCysLeuSerLeuArgThrTrpAspTrpArgPro-HisHisAlaGln 363
 QY 568 GCTAATTAATAAATTTTTGTAGAGACTGGGTCTTACTAGTGTGGCAGCTGTCTT 627
 Db 363 nLeuIlePheCysIle-PheSerArgAsnGlyValLeuProCysTrpProGlyTrpSerA 383
 QY 628 AAATCTCTGGCTTCAAGCAATCTCTTACCTTGGCATCCCAAGTGTGGGATACAGG 687
 Db 393 rgThrProGlyLeu--MetIleHisProLeuProLeuProProlsValLeuGlyLeuGlnAl 402
 QY 688 GTGACCCACCATGTGGCTACTTATTTCTTTACATTCATCTTCCATAGATGTAA 747
 Db 402 a***AlaThrAlaProArgPhePhe----- 411
 QY 748 ATCCACAGACAGGATTACTGCCTATTTCTTCTTCTTTTGTAGACAGAGTCTCAC 807
 Db 412 -----PhePhePhe*****LeuArgGlnSerLeuTh 423
 QY 808 TTCATCACCTCAACCTCCGT-----TCAGCTCACTGCACAACTCTGCCCTCC 852
 Db 423 rLeuSerProArg-LeuGluCysAsnGlyGluIleSerAlaHisCysLysLeuHisLeuP 443
 QY 853 CGGTTCAAGYATCTCTGCTCAAGCTCTGAGTGTGGAATTCACAGCGTGCACC 912
 Db 443 roGlyLeuCysHisSerProAlaSerAlaPheCysValArgGlyThrGlyAlaArg- 462
 QY 913 ACCATGCTTGGCTAATTTTCTATTATTAGCAGAGATGGGTCTTACCATCTGCCAG 972
 Db 463 ThrMetProSer***PhePheValPheLeuValGluMetGlyPheCysHisValGlyGln 482
 QY 973 GCTGCTCAAACTCCTGACCTCAAGTATCTGCTGCTCAGTCTCCAAAGTCTGGA 1032
 Db 483 AlaGlyLeuGluLeuAlaSer***SerThrHisLeuCysLeuProlsCysTrpA 502
 QY 1033 ATTATAGGCTGAGTACCTGCTGCTGGCGGATCTATCTATTTCTTTTATGCTATATC 1092
 Db 502 spTyArgArgGluProLeuHisProAla***** 513
 QY 1093 CCAGATCTAGACAGTGTGCACATATAGTAGTGTCTCAATAAATAATGATGAATGCA 1152
 Db 513 ----- 513
 QY 1153 CAGCTTAGATATAAACTTTCTTTCTTTTAAACAACTTTGACAACTTTGCGAAT 1212
 Db 514 -----PhePhePhePhePhe-----Phe***AspA 523
 QY 1213 AAATCAAACTTGC 1226
 Db 523 rgValSerLeuCys 527

RESULT 5
 A40201
 artifact-warning sequence (translated ALU class A) - human
 C;Species: Homo sapiens (man)
 C;Date: 31-Mar-1992 #sequence_revision 11-Aug-1995 #text_change 19-May-2000
 C;Accession: A40201
 R;Claverie, J.M.
 Personal communication, 1992
 A;Reference number: A40201
 A;Accession: A40201
 A;Molecule type: DNA
 A;Residues: 1-627 <CLA>
 A;Cross-references: UNIPARC:UPI000017CECB
 R;Claverie, J.M.
 Genomics 12, 838-841, 1992

A;Title: Identifying coding exons by similarity search: Alu-derived and other potential
 A;Reference number: A40200; MUID:92241891; PMID:1572661
 A;Contents: annotation
 C;Comment: This "warning" entry is a conceptual translation in all 6 reading frames of c
 in-frame stop codons are shown as 'X'.
 C;Comment: Any significant similarity of a predicted protein sequence to a portion of th

Alignment Scores: 2.06e-26 Length: 627
 Pred. No.: 332.50 Matches: 114
 Score: 332.50 Conservative: 18
 Percent Similarity: 46.6% Mismatches: 86
 Best Local Similarity: 40.3% Indels: 65
 Query Match: 9.4% Gaps: 4
 Db: 4

US-10-607-806-1_copy_7200_9200 (1-2001) x A40201 (1-627)
 QY 397 GTTTGGGTTTTTTTTTTTGTAGACAGAGGCTTTGCTCTGTCAACCCAGGCATGA 456
 Db 320 LeuPheLeuPhePheSer-----GluThrGluSerArgSerValAlaLysAlaGly 336
 QY 457 GCACAGTGTGCAACCATAGTCTACTGAGCTCAACCTCTGAGCTCAAGGATCTGCT 516
 Db 337 ValGlnTrpArgAspLeuGlySerLeuGlnAlaProProGlyPheThrProPheSer 356
 QY 517 GACTCTACCTCCCAAGTAGCTGGAGCTAGAGGCTGCACACCGCTCGCTTAATTA 576
 Db 357 CysLeuSerLeuProSerSerTrpAspTrpArgArgPro-HisTrAlaArgLeuIlePh 376
 QY 577 AAAAAATTTTTGTAGAGACTGGGTCTTACTACGTGGCCAGGCTTGTCTTAAACTCTCTG 636
 Db 376 eCysIle-PheSerArgAspGlyValSerProTrp***ProGlyTrpSerArgSerProA 396
 QY 637 GCTTCAAGCAATCTCTTACCTTGGCATCCCAAAGTGTGGGATTCAGAGGCTGAGCCAC 696
 Db 396 spLeu--ValIleArgProProTrpProProlsValLeuGlyLeuGlnAla----- 412
 QY 697 CATGTCGGCTACTATTTCITTACATTCATCTTCCCAATAGAATGTAAGATCCACAGA 756
 Db 413 -----***AspArgAlaArg** 418
 QY 757 ACAGGGATTACTGCCTATTTTCTTCTTTTGTAGACAGAGTCTCTCACTTCATCAC 816
 Db 418 *****PhePheIlePheCysPheTyThrPhePheProArgSerSerLeuAlaLeuSerPr 438
 QY 817 TCAACTCCGT-----TCAGCTCACTGCAACCTCTGCTCCCGGTTCA 861
 Db 438 oArg-LeuGluCysSerGlyAlaIleSerAlaHisCysLysLeuArgLeuProGlySerA 458
 QY 862 GYGATTCCTCCCTAAGCTCCTGAGTAGCTGGAAATTACAGCGTGCACACCATGCTT 921
 Db 458 rgHisSerProAlaSerAlaSerGlnValAlaGlyThrGlyAlaArg-ThrThrPro 477
 QY 922 GGCTAATTTTGTATTTTGTAGCAGAGATGGGTTTTTACCATTGTTCACAGGCTGGTCTC 981
 Db 478 Gly***PhePheValPheLeuValGluThrGlyPheHisArgGlySerGlnAspGlyLeu 497
 QY 982 AAACCTCCTGACCTCAAGTATCTGCTGCTCAGTCTCCCAAGTGTGGGAATTATAGGC 1041
 Db 498 AspLeuLeuThrSer***--SerAlaArgLeuGlyLeuProlsCysTrpAspTrpArgA 517
 QY 1042 GTGAGTCACTGTGCTGGCGGATTTACTGTCTATTTTCTTTTATTGCTATATCCCATGCT 1101
 Db 517 rgGlu--ThrAlaProGly*****PheLeuPhePheVal----- 530
 QY 1102 AGACAGCTGTCTGACATATAGTAGTGTCTCAATAAATAATGATGAATGCACAGCTAGA 1161
 Db 530 ----- 530
 QY 1162 TATAAACTTTCTTTTCTTTTAAACAACTTTCAGCAACTTTGCAGA 1210
 Db 531 -----PheIlePheArgAspGlyValSerLeuCysArg 542

RESULT 6

A40201
artifact-warning sequence (translated ALU class A) - human
C:Species: Homo sapiens (man)
C:Date: 31-Mar-1992 #sequence_revision 11-Aug-1995 #text_change 19-May-2000
C:Accession: A40201
R:Claverie, J.M.
personal communication, 1992
A:Reference number: A40201
A:Accession: A40201
A:Molecule type: DNA
A:Residues: 1-627 <CLA>
A:Cross-references: UNIPARC:UPI000017C88
R:Claverie, J.M.
Genomics 12, 838-841, 1992

A:Title: Identifying coding exons by similarity search: Alu-derived and other potential
A:Reference number: A40201; PMID:1572661
A:Contents: annotation
C:Comment: This "warning" entry is a conceptual translation in all 6 reading frames of
in-frame stop codons are shown as 'X'.
C:Comment: Any significant similarity of a predicted protein sequence to a portion of th

Alignment Scores:

Pred. No.:	2,68e-25	Length:	627
Score:	322.00	Matches:	104
Percent Similarity:	54.6%	Conservative:	20
Best Local Similarity:	45.8%	Mismatches:	73
Query Match:	9.1%	Indels:	30
DB:	4	Gaps:	2

US-10-607-806-1_COPY_7200_9200 (1-2001) x A40201 (1-627)

QY	1047	ACTCAGCGCTAATTCAGCAGCTTGGGAGACTGAGCAGCAGATCACTTGAGTCTAG	988
Db	5	SerHisAlaCysAsnProSerThrLeuGlyGlnGlyArgGlyLeuThr--ArgSerG	24
QY	987	GAGTTTGAGACAGCCTGGGCAACATGGTAAACCCCATCTCTGTAATAAATACAAAAA	928
Db	24	lyAspArgAspHisProGlyTyHisGlyGluThrProSerLeuLeuLysIleGlnlyI	44
QY	927	TTAGCAGCATGGTGGTGACGCTTGTAAATTCAGCTACTCAGGAGCTTAGCAGGAG	868
Db	44	leSerArgAla--CysGlyArgLeu--SerGlnLeuLeuGlyArgLeuArgGlnGlu	63
QY	867	AATCRCTTGAACCCGGGAGCAGAGTTGCCAGTGAG--CTGAACCGGA	823
Db	64	AsnGlyValAsnProGlyGlyGlyAlaCysSerGluProArgSerArgHisCys-ThrPr	83
QY	822	GTTGAGGTGATGAAGTGAGACTCTGTCTCAAAAAGAGAAAGAAATAGGCAGTAAT	763
Db	83	oAlaLeuAlaThrGluArgAspSerValSerGluLys--	95
QY	762	CCCTGTTCTGTGGATCTTACATTTCTATTGAAAGATGGAATGTAAAGAAATAGTAGCCG	703
Db	96	-----AsnLysAsnLysLys--Lys-----Al	106
QY	702	CACATGGTGGCTCAGCCCTGTAAATCCAGCACTTTGGGATGCCAAGGTAGGAGATTGCT	643
Db	106	aglyArgGlyLeuThrProValIleProAlaLeuTrpGluAlaLysAlaGlyGly--Ser	125
QY	642	TGAAGCCAGGAGTTTAAGCAAGCCTGGCCACGCTAGTAAACCCAGTCTCTACAAAAA	583
Db	126	ArgGlyGlnGluIleGluThrIleLeuAlaThrThrValLysProArgLeuTy--Lys	145
QY	582	ATTTTTPAATAGCCAGCGCTGGTGCGACGCTCGTAGTCCAGCTACTTTGGAGGCT	523
Db	145	styLysLysLeuAlaGlyArgSerAlaGly--AlaCysSerProSerTyLeuGlyGly*	165
QY	522	GAGTTCAGCAGATCCCTTGACCTCAGGAGTTTGGCTGACCTGACCTATGTTGCACCA	463
Db	165	**GlyArgArgMetAla***ThrArgGluAlaGluLeuAlaValSerArgAspArgAlat	185

QY	462	CTGTGCTCATGCTGCTGGTGACAGAGCAAGACCCCTCTCTCTAAACAACAACAAAAAAC	403
Db	185	hrAlaLeuGlnProTrpArgGlnSerGluThrProSerArgLysLysIleLysThrLysA	205
QY	402	CCAAA 398	
Db	205	snLys 206	

RESULT 7
E40201
artifact-warning sequence (translated ALU class E) - human
C:Species: Homo sapiens (man)
C:Date: 31-Mar-1992 #sequence_revision 11-Aug-1995 #text_change 19-May-2000
C:Accession: E40201
R:Claverie, J.M.
personal communication, 1992
A:Reference number: A40201
A:Accession: E40201
A:Molecule type: DNA
A:Residues: 1-597 <CLA>
A:Cross-references: UNIPARC:UPI000017CECC
R:Claverie, J.M.
Genomics 12, 838-841, 1992

A:Title: Identifying coding exons by similarity search: Alu-derived and other potential
A:Reference number: A40201; PMID:1572661
A:Contents: annotation
C:Comment: This "warning" entry is a conceptual translation in all 6 reading frames of
in-frame stop codons are shown as 'X'.
C:Comment: Any significant similarity of a predicted protein sequence to a portion of

Alignment Scores:

Pred. No.:	1,82e-19	Length:	597
Score:	267.00	Matches:	135
Percent Similarity:	42.5%	Conservative:	30
Best Local Similarity:	34.8%	Mismatches:	121
Query Match:	7.6%	Indels:	105
DB:	4	Gaps:	13

US-10-607-806-1_COPY_7200_9200 (1-2001) x E40201 (1-597)

QY	1061	GGCCAGGCACAGTCACTCAGCCCTATATTCAGCAGCTTTGGGAGACTGAGCAGGCAGA	1002
Db	1	GlyGlnValTrp--LeuMetProValIleProAlaLeuTrpGluPheLysAlaGlyArg	20
QY	1001	TC-ATTGAGTTCAGGAGTTTGAGACAGCCTGGGCAACATGTGTAAACCCCATCTCTGC	943
Db	21	IleThr--GlyTrpGluPheGluThrSerLeuThrAsnMetGluLysProHisLeuTy	40
QY	942	TAAATATACAAAAATTAGCAAGCATGGTGTGCAGCTTGTAAATTCAGCTACTCAGG	883
Db	41	***LysTy--LysIleSerTrpAlaTrpTrpArgMetProValIleProAlaThrArg	60
QY	882	AGGCTTAGGCAG--GAGAATCRCTTCAACCCGGGAGGCAGAGTTCAGTGAGCTG--	829
Db	60	luAla-GluThrGlyGluAsnCysLeuAsnProGlySerLysValCysGlyGluIleVal	79
QY	828	-----AACGGAGTTGAGGTGATGA--AGTGAGACTCTGTCTCAAAAAGAAAGAAAGAA	776
Db	80	ProSerTySerGlyLeuGlyAsnLysSerLysThrProSerGlnLysLysLysLysLys	99
QY	775	AATAGGCAGTAATCCCTGTTCTCTGGATCTTACATCTATTGGAAAGATGGAATGTAAG	716
Db	100	***AlaArg-----CysGlySer-----	105
QY	715	AAATAAGTAGCCGACATGGTGTGCTCACCCTCTGAATCCAGCACTTTGGGATGCCAAGG	656
Db	106	-----SerCysLeu--SerGlnArgPheGlySerSerArg	117
QY	655	TA-GGAGGATGTTGTAAGCCAGCAGAGTTTAAGACAAGCCTGGCCCAAGTAGTAGACCCA	597
Db	118	ArgGlyGlySerProGluValGlySerLeuArgProAla--ProThrTrpArgAsnPro	137

```
QY 536 GTCTCTCAAAAAATTTTTTAATTAGCCAGGCGTGTGTGTGACGCTCGTAGTCCGAG 537
   ::::::::::::::::::::
Db 138 lIeSerThrLysAsnIleLys---LeuAlaGlyArgGlyGlyAlaCysLeu***SerGln 156
   ::::::::::::::::::::
QY 536 CTACTTGGAGGCTGAGTCAGCAGATCCCTTGTAGCTCAGGAGGTTGAGGCTGCAGTGAC 477
   ::::::::::::::::::::
Db 157 LeuLeuGlyArgLeuArgGln---GluArgIleAla***ThrGlnGluAlaLysPheAla 175
   ::::::::::::::::::::
QY 476 CTATGGTTGCACACTGTCTCATCTGCTGGGTGCACAGACAA-----435
   ::::::::::::::::::::
Db 176 ValArgLeuCysHisArgThrProAlaIrpAlaThrArgAlaLysLeuArgLeuLysLys 195
   ::::::::::::::::::::
QY 434 -----GACCCCT-----429
   ::::::::::::::::::::
Db 196 LysLys*****ProGlyValValAlaHisAlaCysAsnProSerAlaLeuGlyVal 215
   ::::::::::::::::::::
QY 428 -----GTCTCTAAACAACAACAACAAAAAACCACAA 399
   ::::::::::::::::::::
Db 216 GlnGlyAlaAspHisLeuArgLeuGlyVal***AspGlnLeuAspGlnHisGlyGlu 235
   ::::::::::::::::::::
QY 398 ACTGTCAACCACTTACAAGCTGTGTGAT---CTTGGGCAAAATAGCTCAGCCTTCTGAGCCT 342
   ::::::::::::::::::::
Db 236 ThrProSerLeuLeuLysIle***Asn***LeuGlyValValAlaHisAlaCysAsnPro 255
   ::::::::::::::::::::
QY 341 CAGTTTTCTTATCTGTAAAGTGGAGGATATTGTACCTATCCCATAGTGTGTCTTGAGGA 282
   ::::::::::::::::::::
Db 256 -SerTyrr-----SerGlyGly** 261
   ::::::::::::::::::::
QY 281 TTAAGGAGACACTGCCAGACACAGAAATTAACATTAATCATGCTGTGTGTACTTA 222
   ::::::::::::::::::::
Db 261 *AspArgArgGluLeuLeuGluProArgLysGlnSerLeuArg***AspCysAla--Ile 280
   ::::::::::::::::::::
QY 221 TTATTTTCCCGCGCTACTGAGAACCAACTAGATTATAGTCAAG-----174
   ::::::::::::::::::::
Db 281 ValLeuArgProGlyGlnGlnGlnGlnSerValSerLysLysLysLys*****300
   ::::::::::::::::::::
QY 173 -----GAAGGGATAAACCTACTGCTACAGGTGATTCGCGAGCAGAGCAY 129
   ::::::::::::::::::::
Db 301 PhePhePhePhe***AspGlyValLeuLeuLeu-----ProArgPro 315
   ::::::::::::::::::::
QY 128 GAGTATGAATAGTGTGGGTGTACGKKGTCCAGCAGAAATTTACAGCTGTCAGCTTC 69
   ::::::::::::::::::::
Db 316 GluTyrrAsp-----GlyThrIleSerProGlnThrLeuLeuProGlyPhe 330
   ::::::::::::::::::::
QY 68 TTGGCCTGTCTCR 57
   ::::::::::::::::::::
Db 331 LysGlnPheSer 334
   ::::::::::::::::::::

RESULT 8
I38022
hypothetical protein - human
C:Species: Homo sapiens (man)
C>Date: 01-Nov-1996 #sequence_revision 01-Nov-1996 #text_change 05-Nov-1999
C:Accession: I38022
R:Yang, S.S.; Zhang, K.; Vieira, W.; Taub, J.V.; Zeilstra-Ryalls, J.H.; Somerville, R.L.
A:Title: A human hepatocellular carcinoma 3.0-kilobase DNA sequence transforms both rat
Cancer Res. 50, 5658-5667, 1990
A:Reference number: I38021
A:Accession: I38022
A:Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: DNA
A:Residues: 1-196 <RES>
A:Cross-references: UNIPARC:UPI00011E1FB; EMBL:X55777; NID:g288143; PIDN:CRA39297.1; PH

Alignment Scores:
Pred. No.: 3,43e-19 Length: 196
Score: 264.00 Matches: 61
Percent Similarity: 68.2% Conservative: 12
Best Local Similarity: 57.0% Mismatches: 23
Query Match: 7.5% Indels: 12
DB: 2 Gaps: 3
```

```
US-10-607-806-1_COPY_7200_9200 (1-2001) x I38022 (1-196)
QY 1587 CACTTGAGGTGAGAGTTTCGAGACCAGCTGGTCAATATGGTGAACCCCTATCTCTACTA 1646
   ::::::::::::::::::::
Db 3 HisLeuArgSerGlyValGlnAspTyrProGlyGlnHisGlyLysIleProSerLeuLeu 22
   ::::::::::::::::::::
QY 1647 AAAATACAAAAATPAGCTGGGTGTAGTCATCGCTCTAGTCCCTAGTCCAGCTACTCGGGAGG 1706
   ::::::::::::::::::::
Db 23 LysIleGlnGluLeuAlaGlyHisGlyGlyArgCysLeuGlnSerGlnLeuLeuArgArg 42
   ::::::::::::::::::::
QY 1707 CTGAGGCAAGAGAATTCCTTGAACCTGGGAGGAGAGGTTCCAGTGCAGCCGAGATCCAC 1766
   ::::::::::::::::::::
Db 43 LeuArgGlnGluAsnHisLeuAsnSerGlyGlyArgGlyCysSerGluProLysSerHis 62
   ::::::::::::::::::::
QY 1767 CACTGCATCTCCAGCTGGCGACACAGCAGACTCTATCTCAAAAAATAATAATAATAA 1826
   ::::::::::::::::::::
Db 63 LeuCysIleProAlaIrpValThrGluGlyAspSerValSerLysGln-----AsnLys 80
   ::::::::::::::::::::
QY 1827 ATAAAGGATCGGAG-----AGAAACAATAATAAAGATT-----1861
   ::::::::::::::::::::
Db 81 ThrLysAsnGluGlnHisLeuArgAsnAsnThrLysLysSerAsnSerCysIleIleGly 100
   ::::::::::::::::::::
QY 1862 ---CCTGAAGTAAAGCAGAGA 1879
   ::::::::::::::::::::
Db 101 GlyProGluGlyGluGluLys 107
   ::::::::::::::::::::

RESULT 9
E40201
artifact-warning sequence (translated ALU class E) - human
C:Species: Homo sapiens (man)
C>Date: 31-Mar-1992 #sequence_revision 11-Aug-1995 #text_change 19-May-2000
C:Accession: E40201
R:Claverie, J.M.
A:Title: Identifying coding exons by similarity search: Alu-derived and other potential
Genomics 12, 838-841, 1992
A:Reference number: A40200; MUID:92241891; PMID:1572661
A:Contents: annotation
C:Comment: this "warning" entry is a conceptual translation in all 6 reading frames of
in-frame stop codons are shown as 'X'.
C:Comment: Any significant similarity of a predicted protein sequence to a portion of the
.

Alignment Scores:
Pred. No.: 1,14e-18 Length: 597
Score: 259.50 Matches: 107
Percent Similarity: 40.6% Conservative: 18
Best Local Similarity: 34.7% Mismatches: 75
Query Match: 7.4% Indels: 110
DB: 4 Gaps: 10

US-10-607-806-1_COPY_7200_9200 (1-2001) x E40201 (1-597)
QY 404 TTTTITTTTTTGTGTTGTTGTAGACAGAGGT---CTTGTCTCTGTACCCAGCAGCATGACAC 460
   ::::::::::::::::::::
Db 302 PhePhePhe-----***AspGlyValLeuLeuLeuLeuLeuProArgProGluTyr 317
   ::::::::::::::::::::
QY 461 AGTGGTGCACCAACCATAGTTCACCTGCAGCCTCAACCTCTCAGCTCAAGGATCTGCTG---517
   ::::::::::::::::::::
Db 318 AspGlyThr-----IleSerProGlnThrLeuLeuPro 328
   ::::::::::::::::::::
QY 518 -----ACCTCAGCCTCCCAAGTAGCTGGGAGTACGAGCGTGCAC 556
   ::::::::::::::::::::
Db 329 GlyPheLysGlnPheSerProValSerAlaSerArgValAlaGlyIleThrGlyMetArg 348
   ::::::::::::::::::::
QY 557 CACCACGCTCGCTAATTAATAAAAAATTTTGTAGAGACTGGGTCTTACTACGTTGCC 616
   ::::::::::::::::::::
```


R;Claverie, J.M.
Genomics 12, 838-841, 1992
A;Title: Identifying coding exons by similarity search: Alu-derived and other potential
A;Reference number: A40200; PMID:1572661
C;Contents: annotation
C;Comment: This "warning" entry is a conceptual translation in all 6 reading frames of o
in-frame stop codons are shown as 'X'.
C;Comment: Any significant similarity of a predicted protein sequence to a portion of th

Alignment Scores:
Pred. No.: 1,57e-17 Length: 301
Score: 248.50 Matches: 80
Percent Similarity: 43.7% Conservative: 17
Best Local Similarity: 36.0% Mismatches: 74
Query Match: 7.1% Indels: 52
DB: 4 Gaps: 8

US-10-607-806-1_COPY_7200_9200 (1-2001) x B40201 (1-301)

US-10-607-806-1_COPY_7200_9200 (1-2001) x B40201 (1-301)

Qy 1061 GGCCAGGACAGTGAATCCACCTATAATTCAGCATTGGGAGACTGAGGAGGCGAGA 1002
Db 1 GlyArgAlaTTPLeuThrSerValIleProAlaLeuTrpAspAlaGluValGlyGly 20
Qy 1001 TCACCTGAGGTGAGGAGTTGAGACGACCTGGGACATGGTAACCCCATCTCTGCT 942
Db 21 LeuLeuGluThrArgSerArgProAlaTrpAlaThr***GlnAspProSerSerIle 40
Qy 941 AAAAATACAAAAATTA-----GCCAAGCATGGTGGTGACGCTTGTAA 898
Db 41 LysIleLys***Asn***LeuGly*****AlaGlyHisGlyGly***HisLeu*** 60
Qy 897 TTCAGCTACTCAGGAGCTTAGGACGAGAAATCTTTGAACCCGGGAGGAGGTTGC 838
Db 61 SerGlnHisPheGlyMetLeuLys***GluAspCysLeuArgProGly----- 76
Qy 837 AGTAGCTGAACGGAGGTGAGTGATGAGTGAAGTCTCTCTCAAAAAGAAAGGAAG 778
Db 77 -----ValGlnAspGlnLeuGlyGlnHisSerLysThrProAlaLeu***Lys***Asn 94
Qy 777 AAAATAGGACGTAATCCCTGTTCTCTGGATCTTACATTCTATTGAAAAGATGGAATGAA 718
Db 95 LysIleSer-----TrpVal*****----- 102
Qy 717 AGAAATAGTAGCCGACATGGTGTCTACCCCTGTAATCCAGCATTGGGATGCCAA 658
Db 103 -----GlnGlyMetValAlaAsnIleCysAsnProSerThrLeuGlyCys*** 118
Qy 657 GGTAGGAGATGCTTGAACGAGGAGTTAGACAGCCTGGCCACCACTAGTAGACCC 598
Db 119 SerArgArgIleAla***AspGlnGluPheLysThrSerLeuGlyAsnIleAlaArgPro 138
Qy 597 AGTCTCTACAAA---AAAATT-----TTTTTA 574
Db 139 GlnLeuLysLysAsnLysIleLysLeuAlaGly*****HisProAlaAsnPheIle 158
Qy 573 ATTAGCCAGCGGTGGTGACAGCTCGTAGTCCACGCT-----ACTTGGGAG 526
Db 159 LeuPheLeu***SerTrpGlyLeuAlaMetLeuProLysLeuValLeuAsnSerTrpSer 178
Qy 525 GCTGAGGTGACGAGATCCCTTGGAGTTCAGGAGTTGAGCTGAGTACCTATGTTGCA 466
Db 179 GlnAlaIle-----LeuLeuLeuGlnHisProLysValLeu 190
Qy 465 CCACGTG 460
Db 191 GlyLeu 192

Alignment Scores:
Pred. No.: 1,41e-16 Length: 301
Score: 239.50 Matches: 81
Percent Similarity: 48.6% Conservative: 23
Best Local Similarity: 37.9% Mismatches: 67
Query Match: 6.8% Indels: 44
DB: 4 Gaps: 10

US-10-607-806-1_COPY_7200_9200 (1-2001) x B40201 (1-301)

Qy 467 GCAACCATAGGTCACTGACGACCTCAACCTCCTGAGTCAAGGATCTGCTGACCTCAGCC 526
Db 113 SerThrLeuGlyCys***SerArgArgIleAla***AspGlnGluPheLysThrSerLeu 132
Qy 527 TCCCACTAGCTGGG-----ACTAGGAGCGTGCAC 556
Db 133 GlyAsnIleAlaArgProGlnLeuLysLysAsnLysIleLysLeuAlaGly***** 152
Qy 557 CACCACCCCTGGGTAAATTAATAAATTTTGTGAGAGACTGGGTCTTACCTGTTGGCC 616
Db 153 HisProAlaAsnPheIle-----LeuPheLeu***SerTrpGlyLeuAlaMetLeuPro 170
Qy 617 AGGCTGTCTAAACTCCTCGCTTCAAGCAATCCTCTACCTTGGCATCCCAAGTCTG 676
Db 171 LysLeuValLeuAsnSerTrpSerGlnAlaIleLeuLeuGlnHisProLysValLeu 190
Qy 677 GGATTACAGGGGTGAGCCACCATG---TGGGGCTACTTATTCTTTACATTTCCATCTTTC 733
Db 191 GlyLeuGlnMetLeuAlaThrMetProCys----- 200
Qy 734 CAATAGAATGTAAAGATCCACAGAACAGGGATTACTGCTTATTTCTCTCTTTTCTT 793
Db 201 -----*****ThrGln-----LeuIleLeuPheTyr-----PheTyr 212
Qy 794 AGACAGACTCTCATTCTATCACCCTCAACCTCGTTCAGCTCACTGCAACCTCTGCCTCCC 853
Db 213 ArgAlaGlyValLeuLeuCys---CysProSerTrpSer*****ThrPro 227
Qy 854 GGGTTCAAGYGATTTCTCTGCTAAGCCTCCTGAGTAGCTGGAATTCAAGAGCGTGACCA 913
Db 228 GlyLeuLysGlnSerSerTyrPheSerIleProLysCysTrpAspTyrArgCys***Pro 247
Qy 914 CCATGCTTGGCTAA-----TTTTTTGTATTTTATAGCAGAGATGGGG 954
Db 248 ProCysProAla*****ProSer***PheTyrPheIlePheIle---GluLeuGly 266
Qy 955 TTTTACCATGTTGCCAGGCTGGTCTCAAACTCTGACCTCAAGTGATCTGCTGCTCA 1014
Db 267 SerCysTyrValAlaGlnAlaGlyLeuGluLeuValSerSerAsnProProThrSer 286
Qy 1015 GTCCTCCAAAGTCTGGAATTATAGCGCTGAGTCACTGTGCC 1056
Db 287 AlaSerGlnSerAlaGlyIleThrAspValSerHisHisAla 300

RESULT 13
D40201
artifact-warning sequence (translated ALU class D) - human
C;Species: Homo sapiens (man)
C;Date: 31-Mar-1992 #sequence_revision 11-Aug-1995 #text_change 19-May-2000
C;Accession: D40201
R;Claverie, J.M.
personal communication, 1992
A;Reference number: A40201
A;Cross-references: UNIPARC:UPI000017CEC9

A:Accession: D40201
A:Molecule type: DNA
A:Residues: 1-579 <CLA>
A:Cross-references: UNIPARC:UPI000017CECB
R:Claverie, J.M.
Genomics 12, 838-841, 1992
A:Title: Identifying coding exons by similarity search: Alu-derived and other potential
A:Reference number: A40200; MUID:92241891; PMID:1572661
C:Comments: annotation
C:Comment: This "warning" entry is a conceptual translation in all 6 reading frames of a
in-frame stop codons are shown as 'X'.
C:Comment: Any significant similarity of a predicted protein sequence to a portion of the

Alignment Scores:
Pred. No.: 3,12e-16 Length: 579
Score: 236.50 Matches: 97
Percent Similarity: 41.8% Conservative: 17
Best Local Similarity: 35.5% Mismatches: 77
Query Match: 6.7% Indels: 82
DB: 4 Gaps: 7

US-10-607-806-1_COPY_7200_9200 (1-2001) x D40201 (1-579)

```
QY 397 GTTGGGTT-----TTTTTTTGT-----TGTGT----- 420
      :::::
      339 IleTrpLeuIlePheValSerPhePheCysArgGlyGlyValSerThrCysCysProGly 358
      :::::
      420 -----
      359 CysSer***ThrProGluLeuGluHisThrProAlaMetAlaProLysValLeuGlyLeu 378
      :::::
      421 -----TTAGAGACAGGGTCTTGCTCTGTACCCAG 450
      :::::
      379 GlnAlaCysSerHisAlaTrp*****PheGlnAlaGlySerTrpTyrValAlaGln 398
      :::::
      451 GCATGAGCAGCTGTGTGAACATAGGTCACTGCAGCTCAACCTCTGAGCTCAAGGA 510
      :::::
      399 AlaGlyValGlnArgCysAsnHisGlySerLeu-SerLeuAspPheProGlyGlnVal11 418
      :::::
      511 TCTGTGACCTCAGCTCCCAAGTAGCTGGGACTAGGAGCGTGACACCGCTGGC- 569
      :::::
      418 eleuLeu-ProGlnProPheGlu***LeuGlyLeuGlnAlaTyrAlaThrArgSerGly* 438
      :::::
      570 --TAATAAAAAATTTTCTAGAGACTGGGTCTTACTAGTTGGCAGGCTGTCTT 627
      :::::
      438 **PheLeuTyrLeu-PhePheValGluValGlyPheArgHisValAlaGlnAlaValLeu 457
      :::::
      628 AAATCTCTGGCTTCAAGCAATCTCTACCTTGGCATCCCAAGTGCTGGGATTACAGGG 687
      :::::
      458 GluLeuLeuSerTrpSer-IleHisLeuProTrp-LeuLeuLysCysTrpAspTyrArgA 477
      :::::
      688 GTGAGCCACCATGTGGGCTACTATTCTTTTACATTCATCTTCCAAATGAATGAAG 747
      :::::
      477 rg--AlaAlaMetLeuGly-----
      748 ATCCACAGACAGGGATTACTGCTATTCTTCTCTCTTTTGTAGACAGAGTCTCAC 807
      :::::
      483 -----
      808 TTATCATCACTCAACCTCGTTTCAAGCTCACTG-----CAACCTCTGCTCCCGG 855
      :::::
      491 yMetLeuProArgLeuGluTyrArgGlyAlaIleMetAlaHis**AlaLeuThrSerG1 511
      :::::
      856 GTTCAAGGATCTCTCCCTAAGCTCTCAGTAGCTGGAATACAGCGCTGCACACCC 915
      :::::
      511 nValLys***SerSerCysProSerLeuLeuSerSerTrpAspTyrArgLeuMetLeuPr 531
      :::::
      916 ATGCTTGGCTAATTTT---TTGTATTTTACAGAGATGGGGTTTACCATGTTGCCAG 972
      :::::
      531 oAspLeuAlaAsnPheCysIlePhePheLeu***ArgTrpGlyPheAspMetLeuProAr 551
      :::::
      973 GCTGTCTCAAACTCCTGACCTCAAGTGTATCTGCTGCTGCTCAGTCTCCCAAGTGTCTGA 1032
```

Db 551 gLeuPheLeuAsnSer--***AlaGlyAlaTyrThrCysHisGlySer***SerAlaGly 570
QY 1033 ATTATAGCGCTGAGTCACTGT 1053
Db 571 IleThrGlyValGlnProCys 577
RESULT 14
A46010
X-linked retinopathy protein (C-terminal, clone XEH.8c) - human (fragment)
C:Species: Homo sapiens (man)
C:Date: 21-Sep-1993 #sequence_revision 18-Nov-1994 #text_change 09-Jul-2004
C:Accession: A46010
R:Wong, P.; MacDonald, I.M.; Sood, R.; Smith, C.; Pilon, R.; Tenniswood, M.
Genomics 15, 467-471, 1993
A:Title: Identification and partial characterization of a candidate gene for X-linked
A:Reference number: A46010; MUID:93224131; PMID:8468040
A:Accession: A46010
A:Status: preliminary
A:Molecule type: nucleic acid
A:Residues: 1-100 <WON>
A:Cross-references: UNIPROT:O07826; UNIPARC:UPI000006E23F; GB:S58722; NID:q299470; PID:
A:Note: sequence extracted from NCBI backbone (NCBIN:129339, NCBIPI:129340)

Alignment Scores:
Pred. No.: 7.09e-16 Length: 100
Score: 232.50 Matches: 51
Percent Similarity: 75.3% Conservative: 10
Best Local Similarity: 63.0% Mismatches: 18
Query Match: 6.6% Indels: 2
DB: 2 Gaps: 1

US-10-607-806-1_COPY_7200_9200 (1-2001) x A46010 (1-100)

```
QY 1822 TTTATTATTATTTTGTAGATAGACTCGTGTGTCGCCAGGCTGAGTGCAGTGTGG 1763
      :::::
      2 PhePhePhePhePheGluThrGluSerCysSerValAlaGluAlaGlyValGlnTrpCys 21
      :::::
      1762 GATCTCGCTCAGTCAACCTCTGCTCCCGAGTTCAAGCAATTTCTTGTCTCAGCTC 1703
      :::::
      22 AspLeuGlySerLeuLysSerProPro-ProGlySerSerAspSerProLysAla 41
      :::::
      1702 CCGAGTAGCTGGGACTACAGGCATGATCATCACCCAGCTAATTTTGTATTATTAGT 1643
      :::::
      41 xArgValAlaGlyIleThrGlyMetHisHisThrGlnLeuIlePheValPheLeuVa 61
      :::::
      1642 AGAGATAGGGTTTCAC---CATATTGACAGGCTGTCTCGAACTTCTACCTCAAGTGA 1586
      :::::
      61 lGluThrGlySerHisMetGlnLeuSerAspSerThrLeuValIleThrThrAlaGlnAs 81
      :::::
      1585 T 1585
      81 n 81
      RESULT 15
      S41044
      chromosomal protein - human
      C:Species: Homo sapiens (man)
      C:Date: 19-May-1994 #sequence_revision 10-Nov-1995 #text_change 05-Oct-2004
      C:Accession: S41044
      R:Yeo, J.P.; Alderuccio, F.; Toh, B.H.
      Nature 367, 288-291, 1994
      A:Title: A new chromosomal protein essential for mitotic spindle assembly.
      A:Reference number: S41044; MUID:94166884; PMID:8121495
      A:Accession: S41044
      A:Status: preliminary
      A:Molecule type: mRNA
      A:Residues: 1-418 <YEO>
      A:Cross-references: UNIPROT:P49646; UNIPARC:UPI000013C29A; GB:L26953; NID:g537529; PID:  
C:Superfamily: very very hypothetical protein RNSA-1  
C:Keywords: chromosomal protein
```

Alignment Scores:

Pred. No.:	3,08e-15	Length:	418
Score:	227.00	Matches:	84
Percent Similarity:	40.6%	Conservative:	41
Best Local Similarity:	27.3%	Mismatches:	69
Query Match:	6.4%	Indels:	114
DB:	2	Gaps:	12

US-10-607-806-1_COPY_7200_9200 (1-2001) x S41044 (1-418)

QY	1228	TTCTGCTTTTTCACCTTATCACCCTTGTATGACTTTTTCATATGCTCTCAACCTTTATG	1287
DB	4	PheCysLeuIleLeuPheAsnLeu-----PheAsnPheCys-----LeuIle	17
QY	1288	TTACTGTTTTTTCATGT-----TACTATTTTAGT-----	1317
DB	18	LeuSerPheLeuThrCysIleAlaTyrProValLeuTyrPheSerPheLeuAsnSer	37
QY	1318	-----CACTGAATAATATGG-----	1332
DB	38	GlnThrArgSerLeuLysLeuPheLeuTyrPheLysIleGluIleTrpSerGlnGlyIle	57
QY	1333	CTTAATTTGCTTATACATCCTCTCTGCTCCACTTAGAAGGCCAAATTTACAAATCTG---	1389
DB	58	ValArgPheLeuLeuArgPheGlyLeuThrThrPheLysGluArgPheThrSerLeuIle	77
QY	1389	-----	1389
DB	78	LeuLeuAsnAsnMetHisGlnMetIlePheProMetValLysTyrIleSerLysLeuCys	97
QY	1390	-----ATGAAGCTATGAACCTCTCCCGAGAGAAATACACACAC	1431
DB	98	IlePheHisPheTrpHisLeuValLeuMetAspLeuValProArgGlnArgSer-Ile	117
QY	1432	ACACACACTCACACAGTTTTTTTATGTTTGCACCTAAACAGAAACCTGCATT	1491
DB	117	eIleThrTyrSerLeuValPheAlaIleIleSerGlnLysLysArgGlyIleTyrHi	137
QY	1492	GAGGATGTTTGTTCATATT-----	1510
DB	137	sLysAsnAsnIleArgIleIleLeuPheLeuProGlnAlaHisGlyArgAspPheTyrVa	157
QY	1511	-----AATTAAATACTCAGTTGGGCACAGTCAAGCCTGTAAC	1554
DB	157	lProIleLeuProPheThrGlnSerTyrValAspTrpGlyArgTrpLeu-----	173
QY	1555	CACAGTACTTTGGAAGTCCAAGTGGTGATCCTTGAGTGCAGAAATTCGAGACCAGC	1614
DB	174	-----IleTrpGluAlaLysAlaGlyGluSerLeuGluValArgSerSerArgProAl	191
QY	1615	CTGGTCAATATGTTGAACCTTATCTTACTAAATAACAAATTAAGTTAGCTGGGTAGTG	1674
DB	191	aSerGlnSerArgArgAsnSerValSerThrLysAsnIleLysIleSer-----	207
QY	1675	ATGCATGCTGTA-----	1687
DB	208	-----ProValSerThrLysAsnIleLysIleSerGlnThrTrpTyrLeuPheGlyGl	225
QY	1688	-----GTCCAGCTACTCGGAGGCTGAGGCAAGAAATTCCTTGAACCTGG	1734
DB	225	yValHisLeuLeuValProThrThrArgAspAlaGluAlaGlyGluLeuHisAspPro-G	245
QY	1735	GAGCGAGAGTTGCGAGTCCAGATCCACACTGCTCCAGCTCGGCGGACACAGC	1794
DB	245	lyGlyArgGlyCysAsnGluLeuArgSerCysHisCysThrProAlaTrpValThrSer-	264
QY	1795	GAGACTTCTCAAAAAATAATAATAAAGATCCGAGAGAAACAAACTAA	1854
DB	265	--GluThrValSerLysLys-----LysLysLysLysLysLysLysThrArgVal	282
QY	1855	TAAGATTCCTGAAGTAAGC	1874
DB	282	euthrCysIleAsnAlaSer	288

Search completed: February 8, 2006, 10:03:49
Job time : 87.6544 secs

GenCore version 5.1.7
Copyright (c) 1993 - 2006 Bioceleration Ltd.

OM nucleic - protein search, using frame_plus_n2p model.

Run on: February 8, 2006, 09:18:47 ; Search time 97.1172 Seconds
(without alignments)
2907.338 Million cell updates/sec

Title: US-10-607-806-1_COPY_7200_9200

Perfect score: 3524

Sequence: 1 ctatctcagctgtccctccc.....ckctacataaatatctatta 2001

Scoring table:

BLOSUM62
Xgapop 10.0 , Xgapext 0.5
Ygapop 10.0 , Ygapext 0.5
Fgapop 6.0 , Fgapext 7.0
Delop 6.0 , Delext 7.0

Searched: 2166443 seqs, 705528306 residues

Total number of hits satisfying chosen parameters: 4332886

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Command line parameters:

-MODEL=frame+n2p.model -DEV=xlp
-O=/abs/ABSSWEB spool/US10607806/runat_08022006_091819_26843/app_query.fasta_1
-DB=UniProt -QFMT=fascan -SUFFIX=n2p.rup -MINMATCH=0.1 -LOOPCL=0 -LOOPEXT=0
-UNITS=bits -START=1 -END=1 -MATRIX=blosum62 -TRANS=human40.cdi -LIST=45
-DOALIGN=200 -THR SCORE=pct -THR MAX=100 -THR MIN=0 -ALIGN=15 -MODE=LOCAL
-OUTFMT=pco -NORM=ext -HSPSIZE=500 -MINLEN=0 -MAXLEN=200000000 -HOST=abs02p
-USER=US10607806 @CGN 1.1.694 @runat_08022006_091819_26843 -NCPU=6 -ICPU=3
-NO_WMAP -NEG SCORES=0 -WAIT -DSBLOCK=100 -LONGLOG -DEV TIMEOUT=120
-WARN TIMEOUT=30 -THRAD=1 -XGAPOP=10 -XGAPEXT=0.5 -FGAPOP=6 -FGAPEXT=7
-YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database : UniProt 05.80.*

1: uniprot_sprot.*

2: uniprot_trembl.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	502.5	14.3	591	1 ALU1_HUMAN	P39188 homo sapien
2	498	14.1	591	1 ALU1_HUMAN	P39188 homo sapien
3	479	13.6	593	1 ALU7_HUMAN	P39194 homo sapien
4	463	13.1	593	1 ALU7_HUMAN	P39194 homo sapien
5	443	12.6	593	1 ALU6_HUMAN	P39193 homo sapien
6	438.5	12.4	591	1 ALU8_HUMAN	P39195 homo sapien
7	427.5	12.1	591	1 ALU8_HUMAN	P39195 homo sapien
8	424	12.0	593	1 ALU6_HUMAN	P39193 homo sapien
9	406.5	11.5	587	1 ALU2_HUMAN	P39189 homo sapien
10	397.5	11.3	375	2 O60448 HUMAN	O60448 homo sapien
11	396.5	11.3	239	2 O6ZWA9 HUMAN	O6ZWA9 homo sapien
12	386	11.0	179	2 O6ZTS15_HUMAN	O6ZTS15 homo sapien
13	384.5	10.9	158	2 Q8NAL9_HUMAN	Q8NAL9 homo sapien
14	382.5	10.9	585	1 ALU5_HUMAN	P39192 homo sapien
15	381	10.8	241	2 O6ZTY3_HUMAN	O6ZTY3 homo sapien
16	377.5	10.7	587	1 ALU2_HUMAN	P39189 homo sapien

17	372.5	10.6	585	1	ALU5_HUMAN	P39192 homo sapien
18	372	10.6	200	2	O6ZUK3_HUMAN	O6ZUK3 homo sapien
19	369	10.5	587	1	ALU3_HUMAN	P39190 homo sapien
20	366.5	10.4	156	2	Q8NBH4_HUMAN	Q8NBH4 homo sapien
21	365	10.4	603	1	ALU4_HUMAN	P39191 homo sapien
22	363	10.3	314	2	O6ZTL7_HUMAN	O6ZTL7 homo sapien
23	363	10.3	603	1	ALU4_HUMAN	P39191 homo sapien
24	361	10.2	146	2	O6ZRX6_HUMAN	O6ZRX6 homo sapien
25	360.5	10.2	128	2	O6ZPB2_HUMAN	O6ZPB2 homo sapien
26	350.5	9.9	129	2	O6ZUW8_HUMAN	O6ZUW8 homo sapien
27	344.5	9.8	150	2	O6ZPA0_HUMAN	O6ZPA0 homo sapien
28	344	9.8	123	2	O6ZMQ3_HUMAN	O6ZMQ3 homo sapien
29	342	9.7	129	2	Q8N9K0_HUMAN	Q8N9K0 homo sapien
30	339.5	9.6	176	2	O6ZSR2_HUMAN	O6ZSR2 homo sapien
31	336	9.5	124	2	O6ZSN0_HUMAN	O6ZSN0 homo sapien
32	335.5	9.5	170	2	Q8NAL3_HUMAN	Q8NAL3 homo sapien
33	335.5	9.5	587	1	ALU3_HUMAN	P39190 homo sapien
34	334	9.5	164	2	O6ZUK0_HUMAN	O6ZUK0 homo sapien
35	332.5	9.4	137	2	O6ZS97_HUMAN	O6ZS97 homo sapien
36	331.5	9.4	127	2	O6ZVM2_HUMAN	O6ZVM2 homo sapien
37	329.5	9.4	157	2	Q8N845_HUMAN	Q8N845 homo sapien
38	329	9.3	167	2	O6ZSN3_HUMAN	O6ZSN3 homo sapien
39	328.5	9.3	120	2	O6ZTR3_HUMAN	O6ZTR3 homo sapien
40	328	9.3	140	2	O6ZP99_HUMAN	O6ZP99 homo sapien
41	327.5	9.3	115	2	Q9N083_MACFA	Q9N083 macaca fasc
42	326.5	9.3	131	2	O6ZTD3_HUMAN	O6ZTD3 homo sapien
43	326	9.3	164	2	O6ZTS0_HUMAN	O6ZTS0 homo sapien
44	326	9.3	334	2	O6ZSR97_HUMAN	O6ZSR97 homo sapien
45	325.5	9.2	171	2	Q9H728_HUMAN	Q9H728 homo sapien

ALIGNMENTS

RESULT 1
ALU1_HUMAN
ID ALU1_HUMAN STANDARD; PRT; 591 AA.
AC P39188;
DT 01-FEB-1995 (Rel. 31, Created)
DT 01-FEB-1995 (Rel. 31, Last sequence update)
DT 10-MAY-2005 (Rel. 47, Last annotation update)
DE Alu subfamily J sequence contamination warning entry.
OS Homo sapiens (human)
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini; Hominidae;
OC Homo.
NCBI_TaxID=9606;
RX NCBI
RX NUCLEOTIDE SEQUENCE.
RX MEDLINE=95021759; PubMed=7935834; DOI=10.1038/371752a0;
RA Claverie J.-M., Makalowski W.;
RT "Alu alert".
RL Nature 371:752-752 (1994).
RN [2]
RP CONCEPT.
RX MEDLINE=92241891; PubMed=1572661;
RA Claverie J.-M.;
RT "Identifying coding exons by similarity search: alu-derived and other potentially misleading protein sequences.";
RL Genomics 12:838-841 (1992).
RP ALU FAMILIES CLASSIFICATION.
RX MEDLINE=88333009; PubMed=3138422;
RA Quentin Y.;
RT "The Alu family developed through successive waves of fixation closely connected with primate lineage history.";
RL J. Mol. Evol. 27:194-202 (1988).
RP ALU FAMILIES CLASSIFICATION.
RX MEDLINE=91178815; PubMed=1706781;
RA Jurka J., Milosavljevic A.;
RT "Reconstruction and analysis of human Alu genes.";
RL J. Mol. Evol. 32:105-121 (1991).
CC -|- MISCELLANEOUS: Various analyses indicate that Alu repeats fall

CC into 8 subfamilies (Ref.3 and Ref.4). Therefore, 8 Alu warning
CC consensus sequences have been constituted that contain all six
CC frames conceptual translations of each of these classes of Alu
CC repeats.
CC -!- MISCELLANEOUS: Isolated 'X' indicates the presence of a stop
CC codon, 'XXX' is used to separate the various translation phases.
CC -!- CAUTION: This Alu entry is provided in order to avoid the further
CC pollution of protein sequence databases with Alu-derived amino
CC acid sequences.
CC -!- CAUTION: Alu repetitive sequences are interspersed in human and
CC primate genomes with an average spacing of 4 kb. Some of them are
CC actively transcribed by pol III. Normal transcripts may contain
CC Alu-derived sequences in 5' or 3' untranslated regions. However,
CC cDNA libraries also contain partial and/or rearranged cDNAs
CC ligated with Alu-derived sequence in any orientation. Although Alu
CC elements (especially situated on the complementary strand) have a
CC great potential to create additional/alternative exons,
CC consideration should be given to the possibility that the presence
CC of an Alu in an open reading frame may have resulted from a
CC cloning artifact or may be due to misinterpretation of sequencing
CC data. This point has been overlooked on several occasions, with
CC the consequence of erroneous Alu-derived amino acid sequences
CC being reported.
CC -!- CAUTION: Any significant similarity of a putative protein sequence
CC with an Alu-translated entry must be taken as a warning that a
CC part of Alu repeat may have been artifactually included in the
CC coding nucleotide sequence.
CC -----
CC This Swiss-Prot entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use as long as its content is in no way modified and this statement is not
CC removed.
CC -----
CC EMBL: U04567; -; NOT_ANNOTATED_CDS; Genomic_DNA.
CC KW Hypothetical protein.
CC FT REGION 1 96 Frame-1.
CC FT REGION 100 195 Frame-2.
CC FT REGION 199 294 Frame-3.
CC FT REGION 298 393 Frame-4.
CC FT REGION 397 492 Frame-5.
CC FT REGION 496 591 Frame-6.
CC SQ SEQUENCE 591 AA; 63827 MW; 665D395735519D95 CRC64;

CC Alignment Scores:
CC Pred. No.: 9,46e-42 Length: 591
CC Percent: 502.50 Matches: 129
CC Score Similarity: 48.6% Conservative: 15
CC Best Local Similarity: 43.6% Mismatches: 65
CC Query Match: 14.3% Indels: 87
CC DB: 1 Gaps: 4

CC US-10-607-806-1_COPY_7200_9200 (1-2001) x ALU1_HUMAN (1-591)
CC QY 407 TTTTGTGTTGTTTGTAGACAGGGTCTGTCTGTCCACCGCATGACGACGTGT 466
CC Db 298 PhePheLeu-----ArgGlnGlyLeuAlaLeuSerProArgLeuGluCysSerGly 314
CC QY 467 GCAACATAGTCTACCTGACGCTCAACCTCTGAGCTCAAGGGATCTGTGACCTCAGCC 526
CC Db 315 AlaIleThrAlaHisCysSerLeuAspLeuProGlySerSerAspProProAlaSerAla 334
CC QY 527 TCCCAAGTACTGGGACTAGAGCGTGCACACACGCGCTGGCTAATTAATAAAATTTT 586
CC Db 335 SerArgValAlaGlyThrThrGlyAlaArgHisAlaArgLeuIlePheVal--PheP 354
CC QY 587 TTGTAGAGACTGGTCTTACTAGTGGCCAGCGCTGTCTTAACCTCTGGCTTCAAGCA 646
CC Db 354 heValGluThrGlyPheHisTyxValAlaGlnAlaGlyLeuGluLeuGlySerSera 374
CC QY 647 ATCTCTCTACCTTGGCATCCCAAGTCTCGGATTACAGGGGTGAGCCACCATGTGCGG- 705
CC Db 374 spProProAlaSerAlaSerGlnSerAlaGlyIleThrGlyValSerHisAlaArg* 394

QY 705 ----- 705
Db 394 *****PhePhe***AspArgValSerLeuCysArgProGlyTrpSerAlaValAlaA 414
QY 706 -----CTACTTATTTCTTTTAC 721
Db 414 rgSerArgLeuThrAlaAlaSerThrSerArgAlaGlnAlaIleLeuLeuProGlnProp 434
QY 722 ATTCCATCTTTCCAAATGAGATGTAAAGATCCACAGACAGGATTTACTGCTATTTCTT- 780
Db 434 roGlu**LeuGlyLeuGlnAlaArgAlaThrThrProGly***PheLeuTrpPheLeu* 454
QY 780 ----- 780
Db 454 **ArgArgGlyPheThrMetLeuProArgLeuValSerAsnSerTrpAlaGlnValIleL 474
QY 781 -----C 781
Db 474 euProProArgProProLysValLeuGlyLeuGlnAla**AlaThrAlaProGly**** 494
QY 782 CTTTCTTTTGTGACAGAGTCTCATTCTATCATCTCACTCACTCACTCCGTTCCAGC- 832
Db 494 ****PhePheGluThrGlySerArgSerValAlaGlnAlaGlyVal-GlnTrpArgasp 513
QY 833 -----TCACTGCAACCTCTGCTCCCGGTTCAAGYGATTCTCTGCTAAGCTCTCTG 886
Db 514 HisGlySerLeuGlnProArgProProGlyLeuLysArgSerSerCysLeuSerLeuPro 533
QY 887 AGTAGCTGGAATTACAAAGCGTGCAACACCATGCTTGGCTAATTTTGTATTTTAGCAG 946
Db 534 SerSerTrpAspTyxArgArgAlaProProArgProAlaAsn-PheCysIlePheCysAr 553
QY 947 AGATGGGGTTTATACATGTTGCCAGGCTGCTCAAACTCTGCACTCAAGTGTATGTCG 1006
Db 553 gAspGlyValSerLeuCysCysProGlyTrpSerArgThrProGlyLeuLys***SerSe 573
QY 1007 CTGCTCAGTCTCCCAAGTCTGGAATTATAGGCGTGAG 1046
Db 573 rArgLeuGlyLeuProLysCysTrpAspTyxArgArgGlu 586

CC RESULT 2
CC ALU1_HUMAN
CC ID ALU1_HUMAN STANDARD; PRT; 591 AA.
CC AC P39188;
CC DT 01-FEB-1995 (Rel. 31, Created)
CC DT 01-FEB-1995 (Rel. 31, Last sequence update)
CC DT 10-MAY-2005 (Rel. 47, Last annotation update)
CC DE Alu subfamily J sequence contamination warning entry.
CC OS Homo sapiens (Human).
CC OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
CC OC Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini; Hominidae;
CC OC Homo.
CC OX NCBI_TaxID=9606;
CC RN [1]
CC RP NUCLEOTIDE SEQUENCE.
CC RX MEDLINE=95021758; PubMed=7935834; DOI=10.1038/371752a0;
CC RA Claverie J.-M., Makalowski W.;
CC RT "Alu alert."
CC RL Nature 371:752-752 (1994).
CC RN [2]
CC RP CONCEPT.
CC RX MEDLINE=92241891; PubMed=1572661;
CC RA Claverie J.-M.;
CC RT "Identifying coding exons by similarity search: alu-derived and other
CC RT potentially misleading protein sequences.";
CC RL Genomics 12:838-841 (1992).
CC RN [3]
CC RP ALU FAMILIES CLASSIFICATION.
CC RX MEDLINE=88333009; PubMed=3138422;
CC RA Quentin Y.;
CC RT "The Alu family developed through successive waves of fixation closely
CC RT connected with primate lineage history.";

of an Alu in an open reading frame may have resulted from a cloning artifact or may be due to misinterpretation of sequencing data. This point has been overlooked on several occasions, with the consequence of erroneous Alu-derived amino acid sequences being reported.
 -!- CAUTION: Any significant similarity of a putative protein sequence with an Alu-translated entry must be taken as a warning that a part of Alu repeat may have been artifactually included in the coding nucleotide sequence.

 This Swiss-Prot entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use as long as its content is in no way modified and this statement is not removed.

EMBL; U14574; -; NOT ANNOTATED CDS; Genomic DNA.

REGION	1	96	Frame-1.
FT	100	195	Frame-2.
FT	199	294	Frame-3.
FT	298	393	Frame-4.
FT	397	492	Frame-5.
FT	496	591	Frame-6.
SQ	SEQUENCE 591 AA; 64434 MW; AC8154AD8A6B280 CRC64;		

Alignment Scores:
 Pred. No.: 4,35e-34 Length: 591
 Score: 427.50 Matches: 117
 Percent Similarity: 56.7% Conservative: 10
 Best Local Similarity: 52.2% Mismatches: 59
 Query Match: 12.1% Indels: 38
 DB: 1 Gaps: 4

US-10-607-806-1_COPY_7200_9200 (1-2001) x ALU8_HUMAN (1-591)

QY	420	TTTAGAGACAGGCTTGTCTGTCTGTCACCCAGCAGCATGACACAGTGGTGCACCATAGTCTC	479
Db	398	Phe***AspGlyValSerLeuCyArgProGlyTrpSerAlaValAlaArgSerArgLeu	417
QY	480	ACTGCAGCTCAACCTCTCTGAGTCAAGGATCTGCTGACCTCAGCTCCCAAGTGTG	539
Db	418	ThrAlaThrSerAlaSerArgValGlnAlaLeuLeuProGluProGlu***Leu	437
QY	540	GGACTACGAGCGTGCCACCAACCGCTGCTAATTAATAAATAATTTTGTAGAGACTGG	599
Db	438	GlyLeuGlnAlaArgAlaThrProGly***PheLeuTy-Phe*****-ArgArgG	457
QY	600	GTCTTACTAGTGGCCAGGCTGTCTTAACCTCTGGCTTCAAGCAATCCTCTACCTT	659
Db	457	LysPheThrMetLeuAlaArgLeuValSerAsnSer***ProGlnValIlePheLeu---	475
QY	660	GGCATCCCAAGTGGGATACAGGGTGAGCCACCATGTGCGCTACTTATTTCTTT	719
Db	476	-----Pro-LysCysTrpAspTrpArgArgGluProProArgProAlaSer-AlaArg**	493
QY	720	ACATTCATCTTTTCCCAATAGATGAATGATCCACAGACAGGATTAAGTCTATTTCT	779
Db	493	*****	495
QY	780	TCCTTTCTTTTGTAGACAGAGTCTCACTTCATCCTCACTCCGCTTCAGC-----	832
Db	496	-----PheGluThrGluSerArgSerValAlaGlnAlaGlyVal-GlnTrpArgA	513
QY	833	-----TCATGCAACCTCTGCTCCGGGTTCAAGYGATCTCTGCTCAAGCTCC	884
Db	513	spLeuGlySerLeuGlnProProProGlyPheLeuArgPheSerCysLeuSerLeuP	533
QY	885	TGATAGCTGGAATTAACAGCGTGACCAACCATGCTTGGCTAATTTTGTATTTTAGC	944
Db	533	roSerSerTrpAspTrpArgAlaProProArgProAlaAsn-PheCysIlePheSer	552
QY	945	AGATAGGGGTTTACCATGTTGCCAGGCTGGTCTCAAACTCCTGCACCTCAAGTGTCT	1004

Db	553	ArgAspGlyValSerProCysTrpProGlyTrpSerArgThrProAspLeuArg***Ser	572
QY	1005	GCCTGCCTCAGTCTCCCAAGTCTGGAATATAGCGTGCAGTCACGTGCTGCGCG	1062
Db	573	-----Ser-SerGlnSerAlaGlyIleThrGlyValSerHisArgAlaArgPro	588

RESULT 8

ID	ALU6_HUMAN	STANDARD;	PRT;	593 AA.
AC	P39193;			
DT	01-FEB-1995 (Rel. 31, Created)			
DT	01-FEB-1995 (Rel. 31, Last sequence update)			
DT	10-MAY-2005 (Rel. 47, Last annotation update)			
DE	Alu subfamily SP sequence contamination warning entry.			
OS	Homo sapiens (Human)			
OC	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;			
OC	Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini; Hominidae;			
OC	Homo.			
OX	NCBI_TaxID=9606;			
RN	[1]			
RP	NUCLEOTIDE SEQUENCE..			
RX	MEDLINE=95021758; PubMed=7935834; DOI=10.1038/371752a0;			
RX	Clavexie J.-M., Makalowski W.;			
RT	"Alu alert."			
RL	Nature 371:752-752 (1994).			
RN	[2]			
RP	CONCEPT.			
RX	MEDLINE=92241891; PubMed=1572661;			
RX	Clavexie J.-M.;			
RT	"Identifying coding exons by similarity search: alu-derived and other			
RT	potentially misleading protein sequences.";			
RL	Genomics 12:838-841 (1992).			
RN	[3]			
RP	ALU FAMILIES CLASSIFICATION.			
RX	MEDLINE=88333009; PubMed=3138422;			
RA	Quentin Y.;			
RT	"The Alu family developed through successive waves of fixation closely			
RT	connected with primate lineage history.";			
RL	J. Mol. Evol. 27:194-202 (1988).			
RN	[4]			
RP	ALU FAMILIES CLASSIFICATION.			
RX	MEDLINE=91178815; PubMed=1706781;			
RA	Jurka J., Milosavljevic A.;			
RT	"Reconstruction and analysis of human Alu genes.";			
RL	J. Mol. Evol. 32:105-121 (1991).			
CC	-!- MISCELLANEOUS: Various analyses indicate that Alu repeats fall			
CC	into 8 subfamilies (Ref.3 and Ref.4). Therefore, 8 Alu warning			
CC	consensus sequences have been constituted that contain all six			
CC	frames conceptual translations of each of these classes of Alu			
CC	repeats.			
CC	-!- MISCELLANEOUS: Isolated 'X' indicates the presence of a stop			
CC	codon, 'XXX' is used to separate the various translation phases.			
CC	-!- CAUTION: This Alu entry is provided in order to avoid the further			
CC	pollution of protein sequence databases with Alu-derived amino			
CC	acid sequences.			
CC	-!- CAUTION: Alu repetitive sequences are interspersed in human and			
CC	primate genomes with an average spacing of 4 kb. Some of them are			
CC	actively transcribed by pol III. Normal transcripts may contain			
CC	Alu-derived sequences in 5' or 3' untranslated regions. However,			
CC	cDNA libraries also contain partial and/or rearranged cDNAs			
CC	ligated with Alu-derived sequence in any orientation. Although Alu			
CC	elements (especially situated on the complementary strand) have a			
CC	great potential to create additional/alternative exons,			
CC	consideration should be given to the possibility that the presence			
CC	of an Alu in an open reading frame may have resulted from a			
CC	cloning artifact or may be due to misinterpretation of sequencing			
CC	data. This point has been overlooked on several occasions, with			
CC	the consequence of erroneous Alu-derived amino acid sequences			
CC	being reported.			
CC	-!- CAUTION: Any significant similarity of a putative protein sequence			
CC	with an Alu-translated entry must be taken as a warning that a			
CC	part of Alu repeat may have been artifactually included in the			

```
CC coding nucleotide sequence.
CC -----
CC This Swiss-Prot entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use as long as its content is in no way modified and this statement is not
CC removed.
CC -----
CC EMBL; U14572; -, NOT_ANNOTATED_CDS; Genomic_DNA.
DR InterPro; IPR002218; GIDA.
DR ProDom; PD003738; GIDA; 1.
KW Hypothetical protein.
FT REGION 1 97
FT REGION 101 196
FT REGION 200 295
FT REGION 299 395
FT REGION 399 494
FT REGION 498 593
FT REGION 593 644
SQ SEQUENCE 593 AA; 64640 MW; 136EF344AACD12A2 CRC64;

Alignment Scores:
Pred. No.: 9,93e-34 Length: 593
Score: 424.00 Matches: 113
Percent Similarity: 57.1% Conservative: 16
Best Local Similarity: 50.0% Mismatches: 62
Query Match: 12.0% Indels: 35
DB: 1 Gaps: 3

US-10-607-806-1_COPY_7200_9200 (1-2001) x ALU6_HUMAN (1-593)
QY 1061 GCCAGGCACAGTACTACGCTATATTCAGACACTTTGGAGACTGAGCGAGCAGA 1002
DB 1 GlyArgAlaArgTrpLeuThrProValIleProAlaLeuTrpGluAlaGluAlaGlyGly 20
QY 1001 TCACCTTGAGTTCAGACAGCAGCTGGGCAACATGGTAAACCCCATCTCTCT 942
DB 21 SerProGluValGlySerArgProAla***ProThrTrpArgSerProValIleThr 40
QY 941 AAAATACAAAATTTAGCAGCAGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 882
DB 41 LysAsnThr-LysIleSerArgAlaTyrTrpArgMetProValIleProAlaThrArgG 60
QY 881 GCCTTAGCAGGAGATCTTTGAACCCGGAGGAGGTTGAGTGTGTGTGTGTGTGTGT 827
DB 60 uAlaGluAlaGlyGluSerLeuGluProGlyArgArgArgLeuArg***AlaGluIleAl 80
QY 826 -----CGAGGTTGAGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 777
DB 80 aProLeuHisSerSerLeuGlyAsnLysSerGluThrProSerGlnLys----- 96
QY 776 AAATAGGCAGTAATCCCTGTCTGTGGATCTTACATTTATTTGAAAGATGGAATGTA 717
DB 97 -----Ly 97
QY 716 GAAATAGTACCGCAGATGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 657
DB 97 s*****AlaGlyArgGlyGlySerArgLeu***SerGlnHisPheGlyArgProAr 117
QY 656 GTAGGAGGATGCTTGAAGCCAGCAGGTTTAAAGACAGCCTGGCCCAACCTAGTAGAC 597
DB 117 gArgAlaAspHisLeuArgSerGlyValArgAspGlnProAspGlnHisGlyGluThrPr 137
QY 596 GTCTCTACAAAATAATTTTTTAATTAGCCAGGCGTGTGTGTGTGTGTGTGTGTGTGT 537
DB 137 oSerLeu--LeuLysIleGlnLysLeuAlaGlyArgGlyGlyAlaCysLeu***SerGln 156
QY 536 CTACTTGGGAGGCTGAGGTACAGAGATCCCTGTAGCTCAGAGGTTGAGGCTGAGGTG 477
DB 157 LeuLeuGlyArgLeuArgGlnGluAsnArgLeuAsnProGlyGlyGlyCysGlyGlu 176
QY 476 CTATGTTGACCACTGTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 417
DB 177 ProArgSerArgHisCysThrProAlaThrAlaThrArg-----AlaLysLeu 192
```

```
QY 416 CAACAAAAAAA 405
DB 193 ArgLeuLysLys 196
RESULT 9
ALU2_HUMAN
ID ALU2_HUMAN STANDARD; PRT; 587 AA.
AC P39189;
DT 01-FEB-1995 (Rel. 31, Created)
DT 01-FEB-1995 (Rel. 31, Last sequence update)
DE 10-MAY-2005 (Rel. 47, Last annotation update)
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini; Hominidae;
OC Homo.
OX NCBI_TaxID=9606;
RN [1]
RP NUCLEOTIDE SEQUENCE.
RX MEDLINE=95021758; PubMed=7935834; DOI=10.1038/371752a0;
RA Claverie J.-M., Makalowski W.;
RT "Alu alert."
RL Nature 371:752-752(1994).
RN [2]
RP CONCEPT.
RX MEDLINE=92241891; PubMed=1572661;
RA Claverie J.-M.;
RT "Identifying coding exons by similarity search: alu-derived and other potentially misleading protein sequences."
RL Genomics 12:838-841(1992).
RN [3]
RP ALU FAMILIES CLASSIFICATION.
RX MEDLINE=88333009; PubMed=3138422;
RA Quentin Y.;
RT "The Alu family developed through successive waves of fixation closely connected with primate lineage history."
RL J. Mol. Evol. 27:194-202(1988).
RN [4]
RP ALU FAMILIES CLASSIFICATION.
RX MEDLINE=91178815; PubMed=1706781;
RA Jurka J., Milosavljevic A.;
RT "Reconstruction and analysis of human Alu genes."
RL J. Mol. Evol. 32:105-121(1991).
CC -i- MISCELLANEOUS: Various analyses indicate that Alu repeats fall into 8 subfamilies (Ref.3 and Ref.4). Therefore, 8 Alu warning consensus sequences have been constituted that contain all six frames conceptual translations of each of these classes of Alu repeats.
CC -i- MISCELLANEOUS: Isolated 'X' indicates the presence of a stop codon, 'XX' is used to separate the various translation phases.
CC -i- CAUTION: This Alu entry is provided in order to avoid the further pollution of protein sequence databases with Alu-derived amino acid sequences.
CC -i- CAUTION: Alu repetitive sequences are interspersed in human and primate genomes with an average spacing of 4 kb. Some of them are actively transcribed by pol III. Normal transcripts may contain Alu-derived sequences in 5' or 3' untranslated regions. However, cDNA libraries also contain partial and/or rearranged cDNAs ligated with Alu-derived sequence in any orientation. Although Alu elements (especially situated on the complementary strand) have a great potential to create additional/alternative exons, consideration should be given to the possibility that the presence of an Alu in an open reading frame may have resulted from a cloning artifact or may be due to misinterpretation of sequencing data. This point has been overlooked on several occasions, with the consequence of erroneous Alu-derived amino acid sequences being reported.
CC -i- CAUTION: Any significant similarity of a putative protein sequence with an Alu-translated entry must be taken as a warning that a part of Alu repeat may have been artifactually included in the coding nucleotide sequence.
CC -----
```

CC This Swiss-Prot entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use as long as its content is in no way modified and this statement is not
 CC removed.

DR EMBL; U14568; -, NOT_ANNOTATED_CDS; Genomic_DNA.
 KW Hypothetical protein.
 FT REGION 1 96 Frame-1.
 FT REGION 100 194 Frame-2.
 FT REGION 198 292 Frame-3.
 FT REGION 296 391 Frame-4.
 FT REGION 395 489 Frame-5.
 FT REGION 493 587 Frame-6.
 SQ SEQUENCE 587 AA; 63742 MW; 3EAAB3E3E3929203 CRC64;

Alignment Scores:
 Pred. No.: 6, 08e-32 Length: 587
 Score: 406.50 Matches: 126
 Percent Similarity: 47.2% Conservative: 18
 Best Local Similarity: 41.3% Mismatches: 67
 Query Match: 11.5% Indels: 94
 DB: 1 Gaps: 7

US-10-607-806-1_COPY_7200_9200 (1-2001) x ALU2_HUMAN (1-587)

QY 407 TTTTGTGTTGTTGTAGAGACAGGCTCTGCTGTCTCACCAGGCATGAGCAGTGGT 466
 Db 296 PhePheLeu-----ArgArgSerLeuAlaLeuSerProArgLeuGluCysSergly 312
 QY 467 GCAACATAGTCTACATGACGCTCAACCTCTGACCTCAGGGTCTCTGACCTCAGCC 526
 Db 313 AlalleSerAlaHisCysLeuArgLeuProGlySerArgHisSerProAlaSerAla 332
 QY 527 TCCCAAGTAGCTGGACGACGAGCTGCACACACGCTGCTGCTAAATTAATAAATTTT 586
 Db 333 SerArgValAlaGlyThrThrGlyAlaArgHisAlaAlaArgLeuLeuPheVal--PheL 352
 QY 587 TTGTAGAGATGGGTCTTACTAGTGGCCAGGCTTGCTTAAATCTCTGGCTTCAAGA 646
 Db 352 euValGluThrGlyPheHisArgValSerGlnAspGlyLeuAspLeuLeuThrSer*** 371
 QY 647 ATCTCTCTACCTGGATCCCAAGTCTGGATACAGGGGTGAGCCACCA--TGTCG 703
 Db 372 -SerAlaArgLeuGlyLeuProLysCysTrpAspTyrArgGlyGluProProArgProAl 391
 QY 704 GGTACTATTCTTTACATTCATCTTCCCAATGAATGAAG-----A 748
 Db 391 a*****PhePhe***AspGlyValSer-Leu---CysArgProGlyTrpSerAlaV 410
 QY 749 TCCACAGAACAGGATTAAGTCC----- 771
 Db 410 aAlaArgSerArgLeuThrAlaSerSerAlaSerArgValHisAlaLeuLeuProG 430
 QY 772 -----TATTTCTTCCTT 784
 Db 430 lnProProGlu***LeuGlyLeuGlnAlaProAlaThrProGly***PheLeuTyrP 450
 QY 785 TC----- 786
 Db 450 he*****ArgArgGlyPheThrValLeuAlaArgMetValSerlleSer***ProArgA 470
 QY 786 ----- 786
 Db 470 spProProAlaSerAlaSerGlnSerAlaGlylleThrGlyValSerHisArgAlaArg* 490
 QY 787 -----TTTTTTGAGACAGAGTCTCATTCTACCTCAACCTCGGTTTCAGC----- 832
 Db 490 *****PhePheGluThrGluSerArgSerValAlaGlnAlaGlyVal-GlnTrpArg 509
 QY 833 -----TCACTGCACACCTCTGCCTCCCGGGTTCAAGYGATTCTCTGCTCAAGCTC 883
 Db 510 AspLeuGlySerLeuGlnAlaProProProGlyPheThrProPheSerCysLeuSerLeu 529

QY 884 CTGAGTAGCTGGAAATTAACAAGCGTCACACCATGCTTGCTAAATTTTGTATTTTAG 943
 Db 530 ProSerSerTrpAspTyrArgArgProProArgProAlaAlaSer-PheCysilePheSe 549
 QY 944 CAGAGATGGGGTTTACCATGTTGCCAGCGTGGTCTCAAACTCTCAGCTCAAGTGATC 1003
 Db 549 rArgAspGlyValSerProCys***ProGlyTrpSerArgSerProAspLeu--Valile 568
 QY 1004 TGCTGCTCAGTCTCCCAAGTCTGGAATTATAGCGTGAGTCTACTGTGCTGGC 1060
 Db 569 ArgProProArgProProLysValLeuGlyLeuGlnAla***AlaThrAlaProGly 587

RESULT 10

O60448_HUMAN
 ID O60448 HUMAN PRELIMINARY; PRT; 375 AA.

AC O60448;
 DT 01-AUG-1998 (TrEMBLrel. 07, Created)
 DT 01-AUG-1998 (TrEMBLrel. 07, Last sequence update)
 DT 01-JUN-2003 (TrEMBLrel. 24, Last annotation update)
 DE Neuronal thread protein AD7c-NTP.
 OS Homo sapiens (Human).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini; Hominidae;
 OC Homo.
 OX NCBI_TaxID=9606;
 RN [1]
 RP NUCLEOTIDE SEQUENCE.
 RC TISSUE=Neuronal;
 RX MEDLINE=98064067; PubMed=9399956;
 RA Monte S.M., Ghanbari K., Frey W.H., Beheshti I., Averbach P.,
 HAUSER S.L., Ghanbari H.A., Wands J.R.;
 RT Characterization of the AD7c-NTP cDNA expression in Alzheimer's
 RT disease and measurement of a 41-kD protein in cerebrospinal fluid.;
 RL J. Clin. Invest. 100:3093-3104(1997).
 RN [2]
 RP NUCLEOTIDE SEQUENCE.
 RC TISSUE=Neuronal;
 RA de la Monte S.M., Wands J.R.;
 RL Submitted (JUN-1997) to the EMBL/GenBank/DBJ databases.
 DR EMBL; AF010144; AAC08737.1; -, mRNA.
 DR GO; GO:0005615; C:extracellular space; TAS.
 DR GO; GO:0016021; P:integral to membrane; TAS.
 DR GO; GO:0006915; P:apoptosis; TAS.
 DR GO; GO:0007417; P:central nervous system development; TAS.
 SQ SEQUENCE 375 AA; 41720 MW; 955443950A5BFEFD CRC64;

Alignment Scores:

Pred. No.: 4, 48e-31 Length: 375
 Score: 397.50 Matches: 105
 Percent Similarity: 56.0% Conservative: 12
 Best Local Similarity: 50.2% Mismatches: 60
 Query Match: 11.3% Indels: 32
 DB: 2 Gaps: 4

US-10-607-806-1_COPY_7200_9200 (1-2001) x O60448_HUMAN (1-375)

QY 397 GTTGGGTTTTTTTTTGTGTTGTTTGTAGACAGGCTTGTCTGTCACCCAGGCATGA 456
 Db 194 llePheIlePheIlePheAsnPheLeuArgGlnSerLeuAsnSerValThrGlnAlaGly 213
 QY 457 GCACCTGGTGCACCATAGTCTCAGCTCAGCTCAACTCTCTGAGCTCAAGGGATCTGT 516
 Db 214 ValGlnTrpArgAsnLeuGlySerLeuGlnProLeuProGlyPheIleLeuPheSer 233
 QY 517 GACCTCAGCTCCCAAGTAGTGGGACTACGAGCGTGCACACACGCTGCTAATTA 576
 Db 234 CysProSerLeuLeuSerSerTrpAspTyrArgArg---ProProArgLeuAlaAsnPhe 252
 QY 577 AAAAAATTTTTTGTAGACAGTGGGTCTTACTAGTGGCCAGGCTGTCTTAACTCTCG 636
 Db 253 PheValPheLeuValGlu-MetGlyPheThrMetPheAlaArgLeuIleLeuSer--- 271

Query Match:	11.0%	Indels:	33
DB:	2	Gaps:	6
US-10-607-806-1_COPY_7200_9200 (1-2001) x Q62SI5_HUMAN (1-179)			
QY	433	TCTTGCTCTGCACCCAGGATGACACAGTGGTGCAACCATAGTCTACTGCACGCTCAA	492
DB	3	SerSerSerValSerGlnAlaGlyValGlnTrpLeuAspLeuSerSerLeuGlnProPro	22
QY	493	CCTCTGAGCTCAGGAGTCTGCTGACCTCAGCTCCCAAGTACTGGGACTGACGAGCGT	552
DB	23	ProSerGlyPheLeuArgPheSerCysLeuSerLeuGlnSerSerTrpAspTrpArgHis	42
QY	553	GCACCACACCGCTGGCTTAATTAATAAATTTTTTTT-----GTAGAGACTGGGCT	603
DB	43	AlaProProCysProAlaGly-----PheCysLeuCysValGlnThrGlyPhe	58
QY	604	TACTACGTTGGCAGGCTGTCTTAACTCTGGCTTCAAGCAATCTCTTACCTTGCGCA	663
DB	59	HisHisValGlyGlnAlaSerLeuLeuThrSerSerAspProProAlaLeuAla	78
QY	664	TCCCAAGTGTGGGATTACAGGGGTGAGCCACCATGTGCGGCTACTTATTTTTCAT	723
DB	79	SerGlnIleAlaGlyIleThrGlyMetIleHisHisAlaGlnProLeuLeuPheLeu---	97
QY	724	TCCATCTTTTCCAATAGAAATGTAAGATCCACAGACAGGATTAATGCTCTTCTCT	783
DB	98	-----ArgGlnGlyValThr-----LeuLeuPro	105
QY	784	TTCTTTTGTGACAGAGTCTCATTCTACCTCAACTCGGTTCA-GCTCACTGCAAC	842
DB	106	-----ArgLeuGluCysSerSerAlaIleLeuAlaHisCysIle	118
QY	843	CTCTGCTCCCGGTTCAAGYCATCTCTGCTAAGCTCTGAGTCTGAGTGGGAATTACA	902
DB	119	LeuCysLeuProGlySerSerAspSerProAlaSerAlaSerGlnValAlaGlyThrThr	138
QY	903	AGCGTCACACCATGCTTGGCTAATTTTGTATTTTGTATTTTGTAGCAGAGATGGGT	962
DB	139	GlyVal-----CysProLeuIle-PheValPheValValGluMetGlyPheHisH	155
QY	963	TGTTGCCAGGCTGGTCTCAACTCTCTGACCTCAAGTATGCTGCTGCTCAGTCTCCA	1022
DB	155	eValAla-ArgLeuValPheAsnSerSerProGlnValIleHisProTrpProProL	175
QY	1023	AGTGTGTGGAATTA 1036	
DB	175	ysValLeuArgLeu 179	
RESULT 13			
ID	Q8NAL9_HUMAN	PRELIMINARY;	PRT; 158 AA.
AC	Q8NAL9;		
DT	01-OCT-2002 (TrEMBLrel. 22, Created)		
DT	01-OCT-2002 (TrEMBLrel. 22, Last sequence update)		
DE	Hypothetical protein FLJ35131.		
OS	Homo sapiens (human)		
OC	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;		
OC	Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini; Hominidae;		
OC	Homo.		
OX	NCBI_TaxID=9606;		
RN	[1]		
RP	NUCLEOTIDE SEQUENCE.		
RC	TISSUE=Placenta;		
RX	PubMed=14702039; DOI=10.1038/ngl1285;		
RA	Ota T., Suzuki Y., Nishikawa T., Otsuki T., Sugiyama T., Irie R.,		
RA	Wakamatsu A., Hayaashi K., Sato H., Nagai K., Kimura K., Makita H.,		
RA	Sekine M., Obayashi M., Nishi T., Shibahara T., Tanaka T., Ishii S.,		
RA	Yamamoto J.-I., Saito K., Kawai Y., Isono Y., Nakamura Y.,		
RA	Nagahara K., Murakami K., Yasuda T., Iwayanagi T., Wagatsuma M.,		
RA	Shiratori A., Sudo T., Hosoiri T., Kaku Y., Kodaira H., Kondo H.,		
RA	Sugawara M., Takahashi M., Kanda K., Yokoi T., Furuya T., Kikkawa E.,		


```
QY 957 TTACCATGTTGCCAGGCTGGTCTCAAACTCTGACCTCAAGTATCTGCTGCTCAGT 1016
Db 118 eCyHisIleGlyGlnAlaGlyLeuGluLeuThrSerSerProProThrSerAl 138
QY 1017 CTCCCAAGTCTGGAATATATAGCGTGTGACCTGCTGCTGCTGCTGCTGCTG 1062
Db 138 aSerGlnSerAlaGlyIleThrGlyMetSerHisArgThrTrpPro 153

RESULT 14
ALU5_HUMAN
ID ALU5_HUMAN STANDARD; PRT; 585 AA.
AC P39192;
DT 01-FEB-1995 (Rel. 31, Created)
DT 01-FEB-1995 (Rel. 31, Last sequence update)
DE 10-MAY-2005 (Rel. 47, Last annotation update)
DE Alu subfamily SC sequence contamination warning entry.
OS Homo sapiens (Human)
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini; Hominidae;
OC Homo.
OX NCBI_TaxID=9606;
RN [1]
RP NUCLEOTIDE SEQUENCE.
RX MEDLINE=95021758; PubMed=7935834; DOI=10.1038/371752a0;
RA Claverie J.-M., Makalowski W.;
RT "Alu alert.";
RL Nature 371:752-752(1994).
RN [2]
RP CONCEPT.
RX MEDLINE=92241891; PubMed=1572661;
RA Claverie J.-M.;
RT "Identifying coding exons by similarity search: alu-derived and other
RT potentially misleading protein sequences.";
RL Genomics 12:838-841(1992).
RN [3]
RP ALU FAMILIES CLASSIFICATION.
RX MEDLINE=88333009; PubMed=3138422;
RA Quentin Y.;
RT "The Alu family developed through successive waves of fixation closely
RT connected with primate lineage history.";
RL J. Mol. Evol. 27:194-202(1988).
RN [4]
RP ALU FAMILIES CLASSIFICATION.
RX MEDLINE=91178815; PubMed=1706781;
RA Jurka J., Milosavljevic A.;
RT "Reconstruction and analysis of human Alu genes.";
RL J. Mol. Evol. 32:105-121(1991).
CC -!- MISCELLANEOUS: Various analyses indicate that Alu repeats fall
CC into 8 subfamilies (Ref.3 and Ref.4). Therefore, 8 Alu warning
CC consensus sequences have been constituted that contain all six
CC frames conceptual translations of each of these classes of Alu
CC repeats.
CC -!- MISCELLANEOUS: Isolated 'X' indicates the presence of a stop
CC codon, 'XXX' is used to separate the various translation phases.
CC -!- CAUTION: This Alu entry is provided in order to avoid the further
CC pollution of protein sequence databases with Alu-derived amino
CC acid sequences.
CC -!- CAUTION: Alu repetitive sequences are interspersed in human and
CC primate genomes with an average spacing of 4 kb. Some of them are
CC actively transcribed by pol III. Normal transcripts may contain
CC Alu-derived sequences in 5' or 3' untranslated regions. However,
CC cDNA libraries also contain partial and/or rearranged cDNAs
CC ligated with Alu-derived sequence in any orientation. Although Alu
CC elements (especially situated on the complementary strand) have a
CC great potential to create additional/alternative exons,
CC consideration should be given to the possibility that the presence
CC of an Alu in an open reading frame may have resulted from a
CC cloning artifact or may be due to misinterpretation of sequencing
CC data. This point has been overlooked on several occasions, with
CC the consequence of erroneous Alu-derived amino acid sequences
CC being reported.
CC -!- CAUTION: Any significant similarity of a putative protein sequence
```

```
CC with an Alu-translated entry must be taken as a warning that a
CC part of Alu repeat may have been artifactually included in the
CC coding nucleotide sequence.
CC -----
CC This Swiss-Prot entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use as long as its content is in no way modified and this statement is not
CC removed.
CC -----
DR EMBL; U14571; -, NOT_ANNOTATED_CDS; Genomic_DNA.
KW Hypothetical protein.
FT REGION 1 95 Frame-1.
FT REGION 99 193 Frame-2.
FT REGION 197 291 Frame-3.
FT REGION 295 389 Frame-4.
FT REGION 393 487 Frame-5.
FT REGION 491 585 Frame-6.
SQ SEQUENCE 585 AA; 63994 MW; 46E8C4F493650A7 CRC64;

Alignment Scores:
Pred. No.: 1.72e-29 Length: 585
Score: 382.50 Matches: 108
Percent Similarity: 56.8% Conservative: 18
Best Local Similarity: 48.6% Mismatches: 65
Query Match: 10.9% Indels: 31
DB: 1 Gaps: 2

US-10-607-806-1_COPY_7200_9200 (1-2001) x ALU5_HUMAN (1-585)
QY 1060 GCCAGGCACAGTGACTCACGCTTATATCCAGCAGCTTTGGGAGACTGAGCGGCAGAT 1001
Db 99 AlaglyArgGlyGlySerArgLeu***SerGlnHisPheGlyArgProArgAlaAsp 118
QY 1000 CACTTGAGTTCAGGAGTTTGAGACCCGCTGGGCAACATCGTAAACCCCATCTCTGCTA 941
Db 119 His--GluValLysArgSerArgProSerTrpProThrTrp***AsnProValSerThrL 138
QY 940 AAAATACAAAATAATTAGCCAAGCATGCTGTCGACGCTTCTAATTCAGCTACTCAGGAG 881
Db 138 yAsnThr--LysIleSerTrpAlaTrpTrpArgAlaProValProAlaThrArgGlu 157
QY 880 GCTTAGCGAGGAGAAATCRCTTTGAACCCGGGAGGAGAGGTTTGCAGTGCAGTCAACGAGG 821
Db 158 AlaGluAlaGlyGluSerLeuGluProGlyArgArgGluGln***AlaGlu----- 175
QY 820 TTGAGGTGATGAAGTGAGACTCTGTCTCAAAAAGAAAGAGAGAAATAGCGAGTATATCC 761
Db 176 -----Ile 176
QY 760 CTGTTCTGTGGATCTTACATCTTATTGGAAGATGGAATGTAAGAAATAG-----TA 707
Db 177 AlaProLeuHisSerSerLeuAlaThrGluArgSerValSerLys--Lys***** 196
QY 706 GCGCGCATCATGCGCTCACCCCTTGTATCCAGCAGCTTTGGGATGCCAGGTAGGAGAT 647
Db 196 *ProGlyAlaValAlaHisAlaCysAsnProSerThrLeuGlyGlyArgGlyArgGly 216
QY 646 TGCTTGAGCCAGGAGTTTAAAGACAAGCCCTGGCCACGCTAGTAGTAAGACCCAGTCTTACAA 587
Db 216 eThrArgSerArgAspArg--AspHisProGlyGlnHisGlyGluThrProSerLeu--L 235
QY 586 AAAAAATTTTTTAAATTAGCCAGCGTGTGGTGCAGCTCGTAGTCCAGCTACTTGGGA 527
Db 235 eulysIleGlnLysLeuAlaGlyArgGlyAlaArgLeu***SerGlnLeuLeuGly 255
QY 526 GCGTGAAGTTCAGCAGATCCCTTGCAGCTCAGGAGTTGAGCTGAGTGCAGTATGCTTGC 467
Db 255 rgluArgGlnGluAsnArgLeuAsnProGlyGlyGlyGlyCysSerGluProArgSera 275
QY 466 ACCACTGTGCTATGCTGGGTGACAGACACCCCTGTCTCTTAAACAA 417
Db 275 rghisCysThrProAlaTrpArgGlnSerGluThrProSerGlnLysLys 291
```


GenCore version 5.1.7
Copyright (c) 1993 - 2006 Bioceleration Ltd.

OM nucleic - protein search, using frame_plus_n2p model

Run on: February 8, 2006, 10:22:47; Search time 25.495 Seconds
(without alignments)
3105.549 Million cell updates/sec

Title: US-10-607-806-1-T9182_COPY_8700_9600

Perfect score: 1640

Sequence: 1 tgttcattattataataaataa.....gagctgagcaggagatc 901

Scoring table:

BLOSUM62
Xgapop 10.0, Xgapext 0.5
Ygapop 10.0, Ygapext 0.5
Fgapop 6.0, Fgapext 7.0
Delop 6.0, Delext 7.0

Searched: 2443163 seqs, 439378781 residues

Total number of hits satisfying chosen parameters: 4886326

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Command line parameters:

-MODEL=frame+ n2p model -DEV=xlp
-Q=/abs/ABSSWEB.spool/WARTIN10607806/runat_08022006_092016_27646/app_query.fasta_1
-DB=ABSSWEB -QFMT=fastcan -SURF=n2p.rag -MINMATCH=0.1 -LOOPEXT=0
-UNITS=bits -START=1 -END=1 -MATRIX=blosum62 -TRANS=human40.cdi -LIST=45
-DOCALIGN=200 -THR SCORE=pct -THR MAX=100 -THR MIN=0 -ALIGN=15 -MODE=LOCAL
-OUTFMT=ptc -NORM=ext -HEAPSIZE=500 -MINLEN=0 -MAXLEN=200000000 -HOST=abss02p
-USER=WARTIN10607806 @CGN 1.1 605 @runat_08022006_092016_27646 -NCPU=6 -ICPU=3
-NO MMAP -NEG SCORES=0 -WAIT -DSPBLOCK=100 -LONGLOG -DEV_TIMEOUT=120
-WARN_TIMEOUT=30 -THEADS=1 -XGAPOP=10 -XGAPEXT=0.5 -FGAPOP=6 -FGAPEXT=7
-YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database :

1: Geneseqp21:
2: Geneseqp1980s:
3: Geneseqp1980s:
4: Geneseqp2000s:
5: Geneseqp2001s:
6: Geneseqp2002s:
7: Geneseqp2003as:
8: Geneseqp2003bs:
9: Geneseqp2004s:
9: Geneseqp2005s:

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
C 1	455	27.7	396	AAU30455	Aau30455 Novel hum
C 2	427	26.0	324	AAU29573	Aau29573 Novel hum
C 3	427	26.0	324	AAU29579	Aau29579 Novel hum
C 4	406.5	24.8	397	AAU295913	Aar95913 Neural th
C 5	393.5	24.0	433	ADC87285	Adc87285 Human GPC
C 6	392.5	23.9	241	ABBI1464	Abbi1464 Human neu
C 7	391.5	23.9	179	ADR09211	Adr09211 Human pro
C 8	391	23.8	304	AAU30391	Aau30391 Novel hum
C 9	389.5	23.8	375	AAU01399	Aab01399 Neuron-as

C 10	389.5	23.8	375	5	ABB81538	Abb81538 Neural th
C 11	389.5	23.8	375	5	AAE29142	Aae29142 AD7C-NTP
C 12	389.5	23.8	375	5	ABR63268	AbR63268 AD7C-NTP
C 13	389.5	23.8	375	6	ADA84017	Ada84017 Human POM
C 14	389.5	23.8	375	6	ABU03024	Abu03024 Human neu
C 15	389.5	23.8	375	6	ABU99774	Abu99774 Amino aci
C 16	389.5	23.8	375	6	ABJ19445	Abj19445 AD7C-neur
C 17	389.5	23.8	375	7	ADB37642	AdB37642 Human neu
C 18	389.5	23.8	375	8	ADRI14409	Adri14409 Human NF-
C 19	389.5	23.8	375	9	AEA30020	Aea30020 Human neu
C 20	387.5	23.6	332	5	ADK34410	Adk34410 Novel hum
C 21	387	23.6	411	4	ABG08428	Abg08428 Novel hum
C 22	386	23.5	290	4	ABG21913	Abg21913 Novel hum
C 23	380	23.2	382	4	AAU32610	Aau32610 Novel hum
C 24	380	23.2	382	4	AAU31818	Aau31818 Novel hum
C 25	380	23.2	382	4	AAU32707	Aau32707 Novel hum
C 26	377	23.0	215	4	AAU31513	Aau31513 Novel hum
C 27	376	22.9	302	5	ADK36936	Adk36936 Novel hum
C 28	375	22.9	210	4	AAU31810	Aau31810 Novel hum
C 29	368	22.4	164	8	ADQ67004	Adq67004 Novel hum
C 30	366.5	22.3	156	7	ADB63830	Adb63830 Human pro
C 31	365.5	22.3	381	4	AAU30235	Aau30235 Novel hum
C 32	365	22.3	85	4	AAO11283	Aao11283 Human pol
C 33	364.5	22.2	129	5	AAO16534	Aao16534 Phosphoen
C 34	362.5	22.1	117	5	AAU11247	Aau11247 Human neu
C 35	361	22.0	146	8	ADR09646	Adr09646 Human pro
C 36	360.5	22.0	196	7	ADE09985	Ade09985 Novel pro
C 37	359	21.9	239	8	ADQ65417	Adq65417 Novel hum
C 38	357.5	21.8	673	6	ADA84009	Ada84009 Human POM
C 39	355.5	21.7	175	4	AAU31782	Aau31782 Novel hum
C 40	355.5	21.7	296	4	AAU33304	Aau33304 Novel hum
C 41	354.5	21.6	291	5	ADK36986	Adk36986 Novel hum
C 42	354	21.6	449	5	ABJ01057	Abj01057 Ovary cel
C 43	352	21.5	110	5	ABP41552	Abp41552 Human ova
C 44	351	21.4	241	6	ABP96219	Abp96219 Human nuc
C 45	350.5	21.4	129	8	ADQ65903	Adq65903 Novel hum

ALIGNMENTS

RESULT 1

AAU30455

ID AAU30455 standard; protein; 396 AA.

XX AC AAU30455;

XX AC AAU30455;

DT 18-DEC-2001 (first entry)

XX XX

XX Novel human secreted protein #946.

XX Human; vaccination; gene therapy; nutritional supplement;

KW stem cell proliferation; haematopoiesis; nerve tissue regeneration;

KW immune suppression; immune stimulation; anti-inflammatory; leukaemia.

XX Homo sapiens.

OS Homo sapiens.

XX XX

PN WO200179449-A2.

XX XX

PD 25-OCT-2001.

XX XX

PF 16-APR-2001; 2001WO-US008656.

XX XX

PR 18-APR-2000; 2000US-00552929.

PR 26-JAN-2001; 2001US-00770160.

XX XX

PA (HYSE-) HYSEQ INC.

 XX | Tang YT, Liu C, Drmanac RT; |

QY 605 AAATTTTCCCAATAAAAAAGCTAGTGTATGCTGCCCAAAATCTGGAGGAATATGTAAC 546
 Db 113 rProGluLeuLysGlnSerThrCysLeuSerLeuProLysCysTrp 128
 QY 545 GGTATCTTTGGGTAGTAGGATTAGAG-----TGCTTCCA---CTTTCCATAATAGAGA 495
 Db 129 -----AepTyrArgArgAlaAlaValProGlyLeuPheLeuLeuP 142
 QY 494 TATTATG-----TAGAGTTTGAAGTTG----- 487
 Db 142 hePheLeuArgHisArgCysProThrLeuThrGlnAspGluValGlnTrpCysAspHis 162
 QY 486 -----TAGAGTTTGAAGTTG----- 472
 Db 162 erSerLeuGlnProSerThrLeuArgSerSerIleLeuLeuProGlnProLysVala 182
 QY 471 -----CATTACTTTTGGCTT-----GTTTACTTTTATAACCAAA 438
 Db 182 laGlyThrLysAspMetHisHisThrTrpLeuIlePheIlePheAsnPhel 202
 QY 437 ATAAACAATAGATTACAGTTAAATGCA-----TAATTACGTAATCTC----- 409
 Db 202 euArgGln-----SerLeuAsnSerValThrGlnAlaGlyValGlnTrpArgAsnL 219
 QY 408 -----TTTAACTTTTATACATA----- 391
 Db 219 euGlySerLeuGlnProLeuProGlyPheLysLeuPheSerCysProSerLeuLeuS 239
 QY 390 -----TAATTACGTAATCTC----- 376
 Db 239 erSerTrpAspTyrArgArgProProGlnAlaAsnPheValPheLeuValGlu 259
 QY 375 -----TGCTTAC 369
 Db 259 etGlyPheHisHisValArgGlnValAspAlaArgSerLeuAspLeuValIleCysLeuP 279
 QY 368 CTTC-----GGATCTTATTAGTTTGTCTCTCGATCCTTTATTT 324
 Db 279 roArgProProLysValLeuGlyLeuGlnAspValThrProThrAlaArgProIlePhe 299
 QY 323 ATTTATTTATTTTGGAGTAGAGTCTGCTGTGCTGCCAGGCTGGAGTGGAGTGGTG 264
 Db 299 snPheCysLeu--PheGluMetGluSerHisSerValThrGlnAlaGlyValGlnTrpPr 318
 QY 263 GGATCTCGGCTCACTGCAACTCTGCTCTCCAGGTTCAACAAATCTTCTGCTCAGCCT 204
 Db 318 oAsnLeuGlySerLeuGlnProLeuProGlyLeuLysArgPheSerCysLeuSerLe 338
 QY 203 CCGAGTAGCTGGAGCTACAGGATGATGATGATGATGATGATGATGATGATGATGATGAT 144
 Db 338 uProSerSerTrpAspTyrGlyHisLeu-HisHisThrProLeuIlePheValPheSerL 358
 QY 143 TAGAGATAGGTTTACCATATTGACAGAGCTGGTCTCGAACTTCTCACCTCAAGTGATC 84
 Db 358 euGluAlaGlyPheHisHisIleCysGlnAlaGlyLeuLysLeuLeuThrSerGlyAspP 378
 QY 83 CACCCACCTTGGACTTCCAAAGTACTGTGTTTACAGGCTTGAGT 40
 Db 378 roProAlaSerAlaPheGlnSerAlaGlyIleThrGlyValThr 392
 RESULT 5
 ID ADC87285
 XX ADC87285 standard; protein; 433 AA.
 AC ADC87285;
 XX
 DT 01-JAN-2004 (first entry)
 XX
 DE Human GPCR protein SEQ ID NO:1738.
 XX
 KW human; GPCR; guanosine triphosphate-binding protein coupled receptor;
 gene therapy.

XX Homo sapiens.
 OS
 PN EPI270724-A2.
 XX
 XX 02-JAN-2003.
 XX
 XX 18-JUN-2002; 2002EP-00013517.
 XX
 XX 18-JUN-2001; 2001JP-00246789.
 XX
 XX (NAAD-) NAT INST ADVANCED IND SCI & TECHNOLOGY.
 PA (ADSC-) CENT ADVANCED SCI & TECHNOLOGY INCUBATIO.
 XX
 XX Suwa M, Asai K, Akiyama Y, Aburatani H;
 XX
 XX WPI; 2003-315783/31.
 DR N-PSDB; ADC87284.
 XX
 XX New polynucleotide, useful for preparing a composition for treating a
 PT patient in need of increased or suppressed activity or expression of the
 PT guanosine triphosphate-binding protein coupled receptor.
 PT
 XX
 XX Claim 2; SEQ ID NO 1738; 28pp; English.
 XX
 XX The invention relates to a novel polynucleotide encoding a guanosine
 CC triphosphate-binding protein coupled receptor (GPCR). A polynucleotide of
 CC the invention may have a use in gene therapy. The polynucleotide and
 CC polypeptide are useful for preparing a composition for treating a patient
 CC in need of increased or suppressed activity or expression of the
 CC guanosine triphosphate-binding protein coupled receptor. The protein
 CC sequences shown in ADC85549-ADC87617 represent GPCR's of the invention.
 XX
 XX Sequence 433 AA;
 SQ
 Alignment Scores:
 Pred. No.: 1,86e-39 Length: 433
 Score: 393.50 Matches: 128
 Percent Similarity: 43.8% Conservative: 32
 Best Local Similarity: 35.1% Mismatches: 83
 Query Match: 24.0% Indels: 122
 DB: 7 Gaps: 10
 US-10-607-806-1-T9182_COPY_8700_9600 (1-901) x ADC87285 (1-433)
 QY 899 TACTCTGCTCAGACTCTCTAGTGGAGTACAGTGTGTGCGACATGCCAGCT 840
 Db 100 PheSerCysLeuSerLeuProSerSerTrpAspTyrArgHisValProProArgProAla 119
 QY 839 AATTTTGTATTTTATTAGAGAGGTGTTTCATCATGTGTTCCAGGCTGGTCTCAACT 780
 Db 120 AsnPheCysIlePheSerArgAspGly-PheHisHisValGlyGlnLeuGlyLeuLysLe 139
 QY 779 CCGAGCTCAAGTGTATCCATTTGCTCGGCTCCCAAGTGTGGGATACAGGCGT---G 723
 Db 139 uLeuThrSerSerAspGlnProAlaLeuAlaSerGlnSerAlaGlyIle-AsnArgCysG 159
 QY 722 AGCCACCATGCCAACCTAGTTTCTTACTTTTAAAGCATGTATTTTAAAAATATGGAAC 663
 Db 159 luProProArgProAlaGluValIleSerPhe-SerHisLeuPhe----- 173
 QY 662 ATCTTTAAATGAATACTACATAGACATTTAGAAAAAATAGGTATGCTGCACATGAAA 603
 Db 173 ----- 173
 QY 602 TTTTCCCAATAAAAAAGCTAGTGTATGCTGCCCAAAATCTGGAGGAATATGTAAC 543
 Db 174 -----SerLeuGly 176
 QY 542 TATCTTTGGGTAGTAGGATTATAGATGCTTCCACTTTCCATAATAGATTTA----- 489
 Db 177 PheSerLeuIlePhePheValLeuLeuProGluLysPheProGlnProTyrLeuProIle 196

QY 998 ACTCTGCTCAGACTCTGAGTGGGACTACAGTGTCTGCCACCATGCCAGCTA 839
 Db SerProAlaSerAlaSerArgValAlaGlyThrGlyThrArgHis***AlaArgLeu 53
 QY 838 ATTTTGT-ATTTTATTAGAGGGGTTCATCATGTTCTCAGGCTGTCTCAACT 780
 Db IlePheTrpIlePheSerArgAspGlyValSerProCys***ProGlyTrpSer***Ser 73
 QY 779 CCGGACCTCAAGTATCATTTGCTCGGCTCCCAAAAGTCTGGGATACAGGGTGCAGC 720
 Db ProAspLeu--ValIleArgProProArgLeuProLysCysTrpAspTyArgArgGluP 93
 QY 719 CACATGCCCACTAGTTCTTCTACTTTAAAGCATGTATTTTAAAAAATTATGGAATC 660
 Db roProArgProAla***PhePheValPhe----- 102
 QY 659 CTTAAATGGAATACATACATACACATTTAGAAAAATTAGTATGCTGCACATGGAATTT 600
 Db 102 ----- 102
 QY 599 TCCCAATAAAAAAGCTAGTGTATGCCCAAAATCTGGAGGAATATGTAAGTGTAT 540
 Db 103 -----LeuValGluGlnGlyPheThrMetLeuAlaArgMetValS 116
 QY 539 CTTTGGGTAGTATGATATAGATCCCTTCCACTTTCATATATAGATATTTATGTAGATT 480
 Db 116 exIleSer***ProGlnCysAsp--LeuProAlaSerVal----- 128
 QY 479 TGAAGTTGCATTACTTTTGGCTGTCTTACTTTTATTAACCAAAATAACAATAGATTACA 420
 Db 129 -----SerG 130
 QY 419 GTTAAATGCAATTAAACTTTTATACATATAATTTAGTATCTCTGCTTACCTTCAGGAA 360
 Db 130 naenAlaGlyLeuThrGlyValSerHisCysAlaTrpProCysLeu----- 145
 QY 359 TCTTATTAGTTTGTCTCTCCGACTTTTATTTATTTATTTTATTTTATTTTATTTTATTT 300
 Db 146 -----HisPheCysPhePheGlyPhePheGluMetG 157
 QY 299 GTCTGCTGTCTGCCAGCTGAGTGGAGTGGTGGGATCTCGCTCAGTCACTGCACTCT 240
 Db 157 uSerCysSerValAlaGlnAlaGluValGlnTrpHisAspLeuArgSerLeuGlnAlaPr 177
 QY 239 GCTTCCAGGTTCAGCAATTTCTTCTGCTCCAGCTCCCGAGTGGTGGGACTACAGGCA 180
 Db 177 oProGlyPheThrProPheSerCysLeuSerLeuProGlySerTrpAspTrpArgAr 197
 QY 179 TGCATCACTACACCCAGCTAATTTTGTATTTTATTTATTTATTTATTTATTTATTTATTT 120
 Db 197 gProProArgProAlaAsnPheCysAllePheSerArgAspGlyValSerProCys** 217
 QY 119 ACAGGCTGTCTGCAACTTCTCCTCAAGTATGATCCACCTGAGCTTCCAAAGTA 60
 Db 217 *ProGlyTrpSerArgSerProAspLeu--ValIleArgProProArgProProLysVal 236
 QY 59 CTGTGGTTACAGCT 45
 Db 237 LeuGlyLeuGlnAla 241

US-10-607-806-1-T9182_COPY_8700_9600 (1-901) x ADR09211 (1-179)
 QY 899 TACTCTGCTCAGACTCTGAGTAGTCTGGAGTACAGGTGTGCGACCATGCCAGCT 840
 Db 29 PheSerCysLeuSerLeuGlnSerSerTrpAspTyArgHisAlaProProCysProAla 48
 QY 839 AATTTTGTATTTTATTATAGAGGGTGTTCATCATGTTGTCCAGGCTGTCTCAACT 780
 Db 49 GlyPheCysLeuCysValGluThrGly-PheHisValGlyGlnAlaSerLeuLysLe 68
 QY 779 CCGGACCTCAAGTATCCTATTTGCTCGGCTCCCAAAAGTGTGGATACAGCGTGCAGC 720
 Db 68 uLeuThrSerSerAspProProAlaLeuAlaSerGlnIleAlaGlyIle----- 84
 QY 719 CACCATGCCCACTAGTTTCTTACTTTTAAAGCATGTATTTTAAAAAATTATGGAATC 660

Sequence 179 AA;

Alignment Scores:
 Pred. No.: 2,37e-39 Length: 179
 Score: 391.50 Matches: 104
 Percent Similarity: 41.4% Conservative: 14
 Best Local Similarity: 36.5% Mismatches: 33
 Query Match: 23.9% Indels: 134
 DB: 8 Gaps: 5

RESULT 7
 ADR09211
 ID ADR09211 standard; protein; 179 AA.
 XX
 AC ADR09211;
 XX
 DT 04-NOV-2004 (first entry)
 XX
 DE Human protein useful for treating neurological disease Seq 2717.
 XX
 KW human; oligo-capping method; diagnostic marker; gene therapy;
 KW osteoporosis; neurological disease; Alzheimer's disease;
 KW Parkinson's disease; dementia; short memory; cancer;

KW sense or motor function; emotional reaction; fear response; panic;
 KW osteopathic; neuroprotective; nootropic; antiparkinsonian; cytostatic;
 KW tranquiliser.
 XX Homo sapiens.
 XX EPI447413-A2.
 XX 18-AUG-2004.
 PD 12-FEB-2004; 2004EP-00003145.
 PF 14-FEB-2003; 2003JP-00102207.
 PR 09-MAY-2003; 2003JP-00131452.
 XX (REAS-) RES ASSOC BIOTECHNOLOGY.
 XX Isogai T, Yamamoto J, Nishikawa T, Isono Y, Sugiyama T, Otsuki T;
 PI Wakamatsu A, Ishii S, Nagai K, Irie R;
 XX WPI; 2004-583265/57.
 DR N-PSDB; ADR07255.
 DR
 DR
 XX New 1995 cDNA, useful for treating osteoporosis, neurological diseases,
 PT Alzheimer's diseases, Parkinson's diseases, dementia and various cancers.
 XX
 PS Claim 1; SEQ ID NO 2717; 2686pp; English.
 XX
 CC This invention relates to novel, isolated full length human cDNA
 CC molecules and the encoded proteins thereof. Specifically, it refers to
 CC cDNA clones obtained by an oligo-capping method, where none of these
 CC clones are identical to any known human mRNAs. The present invention
 CC describes an immunoassay to identify agonists and antagonists, as well as
 CC antibodies, antisense molecules and siRNAs that can all be used to bind
 CC to and modulate expression of the cDNA molecules. As such, these
 CC molecules are useful for diagnostic markers or therapeutic targets for
 CC the various diseases or morbid states. In particular, they are useful in
 CC gene therapy for treating osteoporosis, neurological disease, Alzheimer's
 CC disease, Parkinson's disease, dementia, short memory and various cancers,
 CC as well as for maintaining equilibrium of sense or motor function, and
 CC for treating emotional reaction, fear response and panic. Accordingly,
 CC they exhibit osteopathic, neuroprotective, nootropic, antiparkinsonian,
 CC cytotatic and tranquiliser activities. This polypeptide is a protein
 CC encoded by a full length human cDNA sequence of the invention. NOTE: This
 CC sequence is not given in the sequence listing of the specification but
 CC can be obtained on CD-ROM from the European Patent Office, Vienna Sub-
 CC office.

Db	84	-----	84
QY	659	CTTAAATGGAATACTACATAGACATTTAGAAAAATTAGGTATGCTGCACATGGAATTT	600
Db	85	-----	89
QY	599	TCCCAATAAAAAAGCTAGTGTATGCCCAAAAAATCTGGAGGAATATGTAACGTGTTAT	540
Db	89	-----	89
QY	539	CTTTGGGTAGTAGGATTATAGATGCCTTCCACTTTCATAATAGATATTTATGTAGAGTT	480
Db	89	-----	89
QY	479	TGAAGTTGCATTACITTTTTGGCTGTGTTACTTTTATAACCAAAATAAACAATAAGATTACA	420
Db	90	-----	90
QY	419	GTAAAAATGCATTTAAACTTTATTACATATAAATTTACGTATCTCTGCTTACCTTTCAGGAA	360
Db	90	-----	90
QY	359	TCATTATAGTTTGTTCCTCCGATCTCTTCGATCTCTTTATTTATTTATTTTTCAGATAGA	300
Db	91	-----	100
QY	299	GTCTCGCTGTGCGCCAGGCTGGAGTGCAGTGGTGGATCTCGGCTCACTGCAACCTCT	240
Db	100	lyvalThrLeuLeuProArgLeuGluCysSerSerAlaIleLeuAlaHisCysIleLeuC	120
QY	239	GCCTCCAGGTTCAAGCAATTCTCTTGCTCAGCTCCGAGTAGCTGGGACTACAGGCA	180
Db	120	ysLeuProGlySerSerAspSerProAlaSerAlaSerGlnValAlaGlyThrThrGlyV	140
QY	179	TGCATCACTACACCCAGCTAAATTTTGTATTTTGTATTTTAGTAGAGATAGGGTTTCACCATATTG	120
Db	140	al-----CysProLeuIlePheValPheValValGluMetGlyPheHisHisVal-	156
QY	119	ACCAGGCTGCTCGNACTTCTCACTCAAGTGATCCACCACCTGGACTTCCAAAGTA	60
Db	157	AlaArgLeuValPheAsnSerSerProGlnValIleHisProProTrpProProLysVal	176
QY	59	CTGTGGTTA 51	
Db	177	LeuArgLeu 179	
RESULT	8		
AAU30391			
ID	AAU30391	standard; protein; 304 AA.	
XX	AAU30391;		
AC	AAU30391;		
DT	18-DEC-2001	(first entry)	
DE		Novel human secreted protein #882.	
XX			
KW		Human; vaccination; gene therapy; nutritional supplement;	
KW		stem cell proliferation; haematopoiesis; nerve tissue regeneration;	
KW		immune suppression; immune stimulation; anti-inflammatory; leukaemia.	
OS		Homo sapiens.	
XX			
PN	WO200179449-A2.		
XX			
PD	25-OCT-2001.		
XX			
PF	16-APR-2001;	2001WO-US008656.	
XX			
PR	18-APR-2000;	2000US-00552929.	
PR	26-JAN-2001;	2001US-00770160.	
XX			
PA	(HYSE-) HYSEQ INC.		

Db 156 -----PheLeuLysTyrThrLeuAspCysTrp**LySArg 167
 QY 377 TCTGCTTACCTCAGGATCTTATTAGTTTGTCTCTCGATCCTTTAT----- 327
 Db 168 GlnAlaTrpLeu-----TyrTrpTyr***PheGlyThrGluLeuPheGlyCysPro 184
 QY 326 -----TTTATTATTATTATTTTGTGAGATAGAGTCTCGCTGTGCGCCCA 282
 Db 185 **GlyArgGlnThrLeuCysPhe-PhePheLeuArgArgSerLeuAspSerValAlaG 204
 QY 281 GGCTGGAGTCAGTGGGATCTCGGCTCACTGCAACCTCTGCTCCCGAGTTCAAGCA 222
 Db 204 nAlaGlyValGln***CysAspLeuGlySerLeuGlnProLeuProArgPheLys** 224
 QY 221 ATTCTCTTGCTCAGCTCCGAGTCTGGAGTCTGAGCATGAGCATCACTACACCCAGC 162
 Db 224 *PheSerCysLeuSerPheProLeuSerTrpAspTyrArgCysAlaProLeuProAl 244
 QY 161 TAAATTTTGTATTTTATTAGATAGTAG-GGTTTCAACATATTGACCAAGCTGGTCTC-GA 104
 Db 244 aAsnPheCysLeuPheSerArgAspMetGlyPheAlaMetLeuValArgLeuValSerG 264
 QY 103 ACTTCTCACTCAAGTATCACCACCTTGAGCTTCAAAGTACTGTGTACAGCTT 44
 Db 264 uLeuLeuThrSerGlyAspProProAlaSerAlaSerGlnSerAlaGlyIleThrGlyMe 284
 QY 43 GAGTCACTGTGCC 31
 Db 284 tSerTyrArgAla 288

RESULT 9
 AAB01399

ID AAB01399 standard; protein; 375 AA.

XX AAB01399;

AC 20-OCT-2000 (first entry)

DE Neuron-associated protein.

XX Neuron associated protein; NEUPAP; neurological disorder; epilepsy;
 KW ischemic cerebrovascular disease; stroke; cerebral neoplasms;
 KW Alzheimer's disease; Pick's disease; Huntington's disease; dementia;
 KW Parkinson's disease; demyelinating disease; meningitis; prion disease;
 KW kuru; Creutzfeldt-Jakob disease; neurofibromatosis; cerebral palsy;
 KW muscular dystrophy; central nervous system; CNS;
 KW peripheral nervous system; PNS; myopathy; schizophrenia;
 KW actinic keratosis; arteriosclerosis; atherosclerosis; bursitis;
 KW cirrhosis; hepatitis; mixed connective tissue disease; MCTD;
 KW myelofibrosis; paroxysmal nocturnal haemoglobinuria; cancer;
 KW autoimmune disease; inflammation; acquired immunodeficiency syndrome;
 KW AIDS; Addison's disease; adult respiratory distress syndrome; allergy;
 KW ankylosing spondylitis; amyloidosis; anaemia; asthma;
 KW Werner syndrome; trauma; human.

OS Homo sapiens.

XX WO200034477-A2.

PN 15-JUN-2000.

XX 10-DEC-1999; 99WO-US030408.

XX 11-DEC-1998; 98US-00210083.

PR 09-FEB-1999; 99US-0119365P.

PR 16-MAR-1999; 99US-0124687P.

XX (INCY-) INCYTE PHARM INC.

XX Tang YT, Yue H, Baughn MR, Hillman JL, Lal P, Au-Young J;

PI Yang J, Lu DAM, Azimzai Y;

XX

DR WPI; 2000-423423/36.
 XX New human: neuron-associated proteins and polynucleotides encoding them,
 PT useful for diagnosis, treatment and prevention of cell proliferative
 PT disorders including cancer, neuronal and neurological disorders.
 XX
 PS Disclosure; Page 143-144; 145pp; English.
 XX
 CC Human neuron-associated proteins (NEUPAP) can be used for treating or
 CC preventing a disorder associated with decreased expression or activity of
 CC NEUPAP. Antagonists of NEUPAP are useful for treating or preventing
 CC disorder associated with increased expression or activity of NEUPAP. NEUPAP
 CC or their fragments or derivatives are useful for treating neurological
 CC disorder such as epilepsy, ischemic cerebrovascular disease, stroke,
 CC cerebral neoplasms, Alzheimer's disease, Pick's disease, Huntington's
 CC disease, dementia and Parkinson's disease. NEUPAPs are also useful for
 CC treating other demyelinating diseases, bacterial and viral meningitis,
 CC prion diseases including kuru, Creutzfeldt-Jakob disease, nutritional and
 CC metabolic diseases of the nervous system, neurofibromatosis, other
 CC developmental disorders of the central nervous system, cerebral palsy,
 CC neuroskeletal disorders, autonomic nervous system disorders, cranial
 CC nerve disorders, spinal cord diseases, muscular dystrophy and other
 CC neuromuscular disorders, peripheral nervous system disorders, inherited,
 CC metabolic, endocrine, and toxic myopathies, mental disorders including
 CC mood, anxiety and schizophrenic disorders, a cell proliferative disorder
 CC such as actinic keratosis, arteriosclerosis, atherosclerosis, bursitis,
 CC cirrhosis, hepatitis, mixed connective tissue disease (MCTD),
 CC myelofibrosis, paroxysmal nocturnal haemoglobinuria, cancers of the
 CC adrenal gland, bladder, bone, bone marrow, brain, breast, cervix, and an
 CC autoimmune/inflammatory disorder such as acquired immunodeficiency
 CC syndrome (AIDS), Addison's disease, adult respiratory distress syndrome,
 CC allergies, ankylosing spondylitis, amyloidosis, anemia, asthma, Werner
 CC syndrome, complications of cancer, hemodialysis, and extracorporeal
 CC circulation, viral, bacterial, fungal parasitic, protozoal, and
 CC helminthic infections, and trauma. This protein was designated g3002527
 XX

XX Sequence 375 AA;

Alignment Scores:
 Pred. No.: 5.58e-39 Length: 375
 Score: 389.50 Matches: 128
 Percent Similarity: 46.2% Conservative: 40
 Best Local Similarity: 35.2% Mismatches: 88
 Query Match: 23.8% Indels: 108
 DB: 3 Gaps: 15
 US-10-607-806-1-T9182_COPY_8700_9600 (1-901) x AAB01399 (1-375)
 QY 901 GATACCTCTCTGCTCAGCTCTGAGTAGCTGGAGCTACAGGTGCTGCCACCATGCCAG 842
 Db 29 AspSerProAlaSerAlaSerProValAlaGlyIleThrGlyMetCysThrHisAlaArg 48
 QY 841 CTAAAT---TTTGATTTTATTAGAGAGGGTGTTCATCATGTGTCTCCAGGCTGCTCTC 785
 Db 49 LeuIleLeuTyr-PhePheLeuValGluMetGluPheLeuHisValGlyGlnAlaGlyLeu 68
 QY 784 AAATCCCGACCTCAAGTATCCATTTGCTCGGCTCCCAAGTCTGGATACAGGCG 725
 Db 69 GluLeuProThrSerAspAspProSerValSerAlaSerGlnSerAla-ArgTyrArg-T 88
 QY 724 TGAGCCACCATGCCCAA----- 708
 Db 88 hrGlyHisAlaArgLeuCysLeuAlaAsnPheCysGlyArgAsnArgValSerLeuM 108
 QY 707 -----CCTAGTTTCTCT---ACTTTAAAGCAT-----GTATTTTAAATAT 668
 Db 108 etCysProSerTrpSerProGluLeuLysGlnSerThrCysLeuSerLeuProLysCysT 128
 QY 667 GGAACATCTTAAATGGAATACTATACATACATTTAGAAAA----- 627
 Db 128 rpAsp-----TyrArgArgAlaAlaValProGlyLeuP 139
 QY 626 -----ATTAGTATGCTGCG---ACATGGAAATTTTCCCAATAAAAA--- 588

Db	139	heilleLeuPhePheLeuArgHisArgCysProThrLeuThrGlnAspGluValGlnTrpC	159	PR	27-OCT-2000; 2000US-00697590.
Qy	587	-----AAAGCTAGTATGCCCCMAAATCTGGAG-----G	557	XX	(NYMO-) NYMOX PHARM CORP.
Db	159	ysAspHisSerSerLeuGlnProSerThrProGluLeuLysHisProProAlaSerAlas	179	PI	Fitzpatrick J, Averback P, Focht MSS, Bibiano R;
Qy	556	AAATGTAACTGGT-----TATCTTGGGTAGTAGGATTATAGATGC	515	XX	WPI; 2002-507998/54.
Db	179	erGlnValAlaGlyThrLysAspMethHisHisThrTrpLeuLeuPheile-	196	DR	N-PSDB; ABN89470.
Qy	514	CTTCCACTTCCATAATAGATATTATGTAGAGTTTGAAGTTGCTATTACT-	465	XX	New Harilil peptide sequences of the Neural Thread Protein, useful in
Db	197	-----PheilePheAsnPheLeuArgGlnSerLeuAsnSerValThrGlnAlaGlyV	214	PT	therapeutic assays, e.g. as targets for developing drugs for treating
Qy	464	-----TTTGGCTTGTGTACT- 450	450	PT	Alzheimer's disease, ischemia or cerebral infarction, or for diagnosing
Db	214	alGlnTrpArgAsnLeuGlySerLeuGlnProLeuProGlyPheLysLeuPheSerC	234	XX	these diseases.
Qy	449	-----TTTA 446	446	PS	Example 1; Fig 1; 53pp; English.
Db	234	ysProSerLeuLeuSerSerTrpAspTyrArgArgProProArgLeuAlaAsnPheV	254	XX	The present invention describes a neural thread protein (NTP) peptide
Qy	445	TAACCAAAATAACAATAAGATTACAGTTAAATGCAATTTTAACTTATTACA-----T	392	CC	having an amino acid sequence selected from ABB81511 to ABB81529 and
Db	254	alPheLeuValGluMetGlyPheThrMetPheAlaArgLeuLeuLeuSerGlyProc	274	CC	their homologues, which are referred collectively as Harilil peptides (I).
Qy	391	ATAATTCAGTATCTCGTTACCTTCAGGAATCTTATTAGTTTGTTCCTCCGATCC	332	CC	(I) have neuroprotective, nootropic, vasotropic and cerebroprotective
Db	274	ysAspLeuProAlaSerAlaSerGlnSerAlaGly-IleThrGlyValSerHisAla	293	CC	activities, and can be used in peptide therapy. The Harilil peptide
Qy	331	TTTATTTTATTATTAT--TTTTCAGATAGAGTCTCGTCTGTGCGCCAGGCTGGA	275	CC	sequences can be used as analogues for NTP in therapeutic or diagnostic
Db	294	ArgLeuLeuPheAsnPheCysLeuPheGluMetGlnSerHisSerValThrGlnAlaGly	313	CC	assays by replacing NTP with the peptide in such an assay. The Harilil
Qy	274	GTGAGTGGTGGGATCTCGGTCCTCACTCGCTCCAGGTTCAAGCAATTTCTCT	215	CC	peptides are also useful as a trap material in a diagnostic or
Db	314	ValGlnTrpProAsnLeuGlySerLeuGlnProLeuProGlyLeuLysArgPheSer	333	CC	therapeutic assay. Therefore, the Harilil peptides are useful in binding
Qy	214	TGCCTCAGCTCCGAGTAGCTGGAGTACAGGATGATCATCATACCCAGCTAATTTT	155	CC	assays, protein and antibody purification, therapeutics or diagnostics.
Db	334	CysLeuSerLeuProSerSerTrpAspTyrGlyHisLeuProProHisProAlaAsnPhe	353	CC	In particular, the peptides are also useful for diagnosing Alzheimer's
Qy	154	TGTATTTTGTAGTACAGATAGGTTTTCACCATATTGACAGGCTGTCTCGAATCTTCAC	95	CC	disease, Down's syndrome, neuroectodermal tumours, astrocytomas,
Db	354	CyslePheleleArgGlyGlyValSerProTyrLeuSerGlyTrpSerGlnThrProAsp	373	CC	glioblastomas, hypoxia, ischaemia or cerebral infarction. The peptides
Qy	94	CTCAAG 89	89	CC	are also useful as targets for drug development for the treatment of
Db	374	LeuArg 375	375	CC	these diseases. The present sequence represents a neural thread protein
RESULT 10				XX	given in the exemplification of the present invention
ABB81538				XX	Sequence 375 AA;
ID	ABB81538	standard; protein; 375 AA.		XX	
AC	ABB81538;			XX	
XX	02-SEP-2002	(first entry)		XX	
DT	Neural thread protein (NTP).			XX	
DE	Neural thread protein; NTP; Harilil peptide; Alzheimer's disease;			XX	
KW	Down's syndrome; neuroectodermal tumour; astrocytoma; glioblastoma;			XX	
KW	hypoxia; ischaemia; cerebral infarction.			XX	
XX	Homo sapiens.			XX	
XX	WO200234915-A2.			XX	
PN	02-MAY-2002.			XX	
XX	25-OCT-2001; 2001WO-US042813.			XX	
FD				XX	
XX				XX	
PF				XX	
XX				XX	

Alignment Scores:					
Pred. No.:	5.58e-39	Length:	375		
Score:	389.50	Matches:	128		
Percent Similarity:	46.2%	Conservative:	40		
Best Local Similarity:	35.2%	Mismatches:	88		
Query Match:	23.8%	Indels:	108		
DB:	5	Gaps:	15		

US-10-607-806-1-T9182_COPY_8700_9600 (1-901) x ABB81538 (1-375)

Qy	901	GATACCTCTCCCTCAGACTCTCAGTAGCTGGAGCTACAGGCTTGCACCATGCCAG	842		
Db	29	AspSerProAlaSerAlaSerProValAlaGlyIleThrGlyMetCysThrHisAlaArg	48		
Qy	841	CTAATT---TTTGTATTTTATTAGAGGGTGTTCATCATCTTGTCCAGGCTGGTCTC	785		
Db	49	LeuileuTyrPhePheLeuValGluMetGluPheLeuHisValGlyGlnAlaGlyLeu	68		
Qy	784	AAACTCCGGACCTCAAGTATCATTTGGCTCCGCTCCCAAGTGTGGGATACAGCG	725		
Db	69	GluLeuProThrSerAspAspProSerValSerAlaSerGlnSerAla-ArgTyrArg-T	88		
Qy	724	TGAGCCACCATGCCAA-----	708		
Db	88	hrGlyHisHisAlaArgLeuCysLeuAlaAsnPheCysGlyArgAsnArgValSerLeuM	108		
Qy	707	-----CCTAGTTTCTCT---ACTTTAAAGCAT-----GTATTTTAAAAATAT	668		
Db	108	etCysProSerTrpSerProGluLeuLysGlnSerThrCysLeuSerLeuProLysCysT	128		
Qy	667	GGAACATCTTAAATGGAATACATACATAGACATTGAAAA-----	627		
Db	128	rpAsp-----TyrArgArgAlaAlaValProGlyLeuP	139		
Qy	626	-----ATTAGGTATGCTGC---ACATGGAAATTTTCCCAATAAAA-----	588		
Db	139	heileLeuPhePheLeuArgHisArgCysProThrLeuThrGlnAspGluValGlnTrpC	159		

QY 587 -----AAAGCTAGTGTATGCCCCAAAAATCTGGAG-----G 557
 Db 159 ysAspHisSerSerLeuGlnProSerThrProGluLeuLysHisProProAlaSerAla 179
 QY 556 AATATGTAACCTGGT-----TATCTTTGGTAGTAGGATTATAGATGC 515
 Db 179 erGlnValAlaGlyThrLysAspMetHisIstYrThrTrpLeuPheile----- 196
 QY 514 CTTCACCTTTCCATAATAGATATTTATGTAGAGTTTGAAGTTGCATTACT----- 465
 Db 197 -----PheIlePheAsnPheLeuArgGlnSerLeuAsnSerValThrGlnAlaGlyV 214
 QY 464 -----TTCGGCTGTGTTACT- 450
 Db 214 alGlnTrpArgAsnLeuGlySerLeuGlnProLeuProGlyPheLysLeuPheSerC 234
 QY 449 -----TTTTA 446
 Db 234 ysProSerLeuLeuSerSerTrpAspTyrArgArgProProArgLeuAlaAsnPhePheV 254
 QY 445 TAACCAAAATAACAATAAGATTACAGTTAAATGCACTTTAACTTTATTACA-----T 392
 Db 254 alPheLeuValGluMetGlyPheThrMetPheAlaArgLeuLeuLeuSerGlyProC 274
 QY 391 ATAATTACGTATCTCTGCTTACCTTCAGGAATCTTATTAGTTTGTCTCCGATCC 332
 Db 274 ysAspLeuProAlaSerAlaSerGlnSerAlaGly-IleThrGlyValSerHisAla 293
 QY 331 TTTATTTTATTATTAT-----TTTTTGATAGATGCTCGCTGTGCGCCAGCTGGA 275
 Db 294 ArgLeuIlePheAsnPheCysLeuPheGluMetGluSerHisSerValThrGlnAlaGly 313
 QY 274 GTGCAGTGTGGGATCTCGCTCACTCAACCTCTGCTCCAGTTTCAAGCAATCTCT 215
 Db 314 ValGlnTrpProAsnLeuGlySerLeuGlnProLeuProGlyLeuLysArgPheSer 333
 QY 214 TGCTCAGCTCCCGAGTACTGGGACTCAGGATGATCATCTACACCCAGCTAATTTT 155
 Db 334 CysLeuSerLeuProSerSerTrpAspTyrGlyHisLeuProHisProAlaAsnPhe 353
 QY 154 TGTATTTTATAGATAGAGTGGTTTACCATATTTGACCGCTGGCTCGAATCTCTCAC 95
 Db 354 CysIlePheIleArgGlyValSerProTyrLeuSerGlyTrpSerGlnThrProAsp 373
 QY 94 CTCAAG 89
 Db 374 LeuArg 375
 RESULT 11
 ID AAE29142
 XX AAE29142 standard; protein; 375 AA.
 AC AAE29142;
 XX
 DT 27-JAN-2003 (first entry)
 XX
 DE AD7c-NTP protein.
 XX
 KW Neural thread protein; NTP; hyperplasia; hypertrophy; arteriosclerosis;
 KW haemorrhoid; gene therapy; tumour; vascular disease; atherosclerosis;
 KW inflammatory disease; nutritional deficiency disease; genetic disease;
 KW autoimmune disease; metabolic disease; traumatic disease; intoxication;
 KW infectious disease; congenital malformation; enzyme deficiency disease;
 KW amyloid disease; fibrosis disease; storage disease; radiation disease;
 KW poisoning; environmental disease; endocrine disease; protein therapy;
 KW degenerative disease; mechanical disease; AD7c-NTP protein.
 XX
 OS Unidentified.
 XX
 XX
 PN WO200274323-A2.
 XX
 PD 26-SEP-2002.
 XX

PF 08-MAR-2002; 2002MO-IB001959.
 XX
 PR 08-MAR-2001; 2001US-0273957P.
 XX
 PA (AVER/) AVERBACK P.
 XX
 PI AVerback P;
 XX
 XX WPI; 2002-759864/82.
 DR N-PSDB; AAD46671.
 XX
 XX
 PT Treating a condition in a patient requiring removal or destruction of
 PT cells, such as a benign or malignant tumor of a tissue or an inflammatory
 PT disease, comprises administering a neural thread protein (NTP) or a NTP
 PT gene to a mammal.
 XX
 PS Example 2; Fig 1; 70pp; English.
 XX
 CC The invention relates to a method for treating a condition in a patient
 CC requiring removal or destruction of cells. The method involves
 CC administering to a mammal a neural thread protein (NTP), or administering
 CC to a tumour or other target cell a NTP gene, where the expression of the
 CC NTP gene is induced resulting in expression of the NTP protein. The
 CC method and NTP are useful for treating a condition in a patient requiring
 CC removal or destruction of cells, such as a benign or malignant tumour of
 CC a tissue, a hyperplasia, hypertrophy, or overgrowth of a tissue,
 CC preferably tonsillar hypertrophy or prostatic hyperplasia, a virally,
 CC bacterially, or parasitically altered tissue, or a malformation of a
 CC tissue. Other conditions include a cosmetic modification to a tissue,
 CC such as removal of unwanted facial hair, warts or unwanted fatty tissue,
 CC a vascular disease, particularly atherosclerosis or arteriosclerosis,
 CC haemorrhoids, or varicose veins, an inflammatory disease, autoimmune
 CC disease, metabolic disease, hereditary/genetic disease, traumatic disease
 CC or physical injury, nutritional deficiency disease, infectious disease,
 CC congenital malformation, amyloid disease, fibrosis disease, storage
 CC disease, enzyme deficiency disease, poisoning, intoxication, degenerative
 CC disease, radiation disease, environmental disease, endocrine disease or
 CC mechanical disease. The invention is useful in protein therapy and gene
 CC therapy. The present sequence is AD7c-NTP protein
 XX
 SQ Sequence 375 AA;
 Alignment Scores:
 Pred. No.: 5.58e-39 Length: 375
 Score: 389.50 Matches: 128
 Percent Similarity: 46.2% Conservative: 40
 Best Local Similarity: 35.2% Mismatches: 88
 Query Match: 23.8% Indels: 108
 DB: Gaps: 15
 US-10-607-806-1-T9182_COPY_8700_9600 (1-901) x AAE29142 (1-375)
 QY 901 GATATCTCTGCTCAGACTCTCTGAGTAGTGGAGCTACAGGTGCTTGGCCACCATGCCAG 842
 Db 29 AspSerProAlaSerAlaSerProValAlaGlyIleThrGlyMetCysThrHisAlaArg 48
 QY 841 CTAATTT---TTTGATTTTATTAGAGAGGGTGTTCATCATGTCTCCAGGCTGCTTC 785
 Db 49 LeuIleLeuTyrPhePheLeuValGluMetGluPheLeuHisValGlyGlnAlaGlyLeu 68
 QY 784 AAATCCCGACCTCAAGTATCATTTGCTCGGCTCCAAAGTCTGGGATACAGGCG 725
 Db 69 GluLeuProThrSerAspSerProSerValSerAlaSerGlnSerAla-ArgTyrArg-T 88
 QY 724 TGAGCCACCATGCCAA----- 708
 Db 88 hrGlyHisHisAlaArgLeuCysLeuAlaAsnPheCysGlyArgAenArgValSerLeuM 108
 QY 707 -----CCTAGTTTCTCT---ACTTTAAAGCAT-----GTATTTTAAATTTAT 568
 Db 108 etCysProSerTrpSerProGluLeuLysGlnSerThrCysLeuSerLeuProLysCysT 128
 QY 667 GGAACATCTTAAATGAATACTACATAGACATTTAGAAAA----- 627

Db	128	rpAsp			----	TyArgArgAlaAlaValProGlyLeuP	139
QY	626		----	ATTAGGTATGCTGC	----	ACATGGAATTTTCCCAATAAAA	588
Db	139	heLeuLeuPheLeu	argHisArgCysProThrLeuThrGlnAspGluValGlnTrpC	159			
QY	587	----	AAAGCTAGTGTATGCCCAAAAATCTGGAG	----			G 557
Db	159	ysAspHisSerSerLeuGlnProSerThrProGluLeuLysHisProProAlaSerAlas	179				
QY	556	ATATGTAACTGGT	----	TATCTTTGGGTAGTAGGTTATAGATCC	515		
Db	179	exGlnValAlaGlyThrLysAspMethHisHisIstyrThrTrpLeuIlePheIle	196				
QY	514	CTTCCACTTCCATAATAGATATTATGTAGAGTTTGAAGTTCGATTACT	465				
Db	197	-----PheIlePheAsnPheLeuArgGlnSerLeuAsnSerValThrGlnAlaGlyVal	214				
QY	464		----	TTTGGCTGTTTACT	450		
Db	214	alGlnTrpArgAsnLeuGlySerLeuGlnProLeuProGlyPheLysLeuPheSerC	234				
QY	449		----	TTTAACTTTATTACA	392		
Db	234	ysProSerLeuLeuSerSerTrpAspTyrArgArgProProArgLeuAlaAsnPhePheV	254				
QY	445	TAACCAAAATAACAATAAGATTACAGTTAAATGCATTTAACTTTATTACA	392				
Db	254	alPheLeuValGluMetGlyPheThrMetPheAlaArgLeuLeuLeuSerGlyProC	274				
QY	391	ATAATTTACGTACTCTGCTTACCTTCAGGAATCTTATTAGTTTGTGTTCTCCGATCC	332				
Db	274	ysAspLeuProAlaSerAlaSerGlnSerAlaGlyIleThrGlyValSerHisAla	293				
QY	331	TTTATTTTATTATTAT	----	TTTTTTCAGATAGAGTCTCGCTGTGTGCGCCAGGCTGGA	275		
Db	294	ArgLeuIlePheAsnPheCysLeuPheGluMetGluSerHisSerValThrGlnAlaGly	313				
QY	274	GTCAGTGTGGGATCTGGCTCACTGCAACTCTGCTCCAGTCCAGGTTCAAGCAATTTCTCT	215				
Db	314	ValGlnTrpProAsnLeuGlySerLeuGlnProLeuProGlyLeuLysArgPheSer	333				
QY	214	TGCTCAGCTCCCGAGTAGCTGGGACTACAGGCATGCATCACACCCAGCTAATTTT	155				
Db	334	CysLeuSerLeuProSerSerTrpAspTyrGlyHisLeuProProHisProAlaAsnPhe	353				
QY	154	TGPTATTTTAGTAGATAGGTTTCCACATATTGACACGCTGGTCTCGAATCTCTCAC	95				
Db	354	CysIlePheIleArgGlyGlyValSerProtyrLeuSerGlyTrpSerGlnThrProAsp	373				
QY	94	CTCAAG	89				
Db	374	LeuArg	375				
RESULT 12							
ABR63268							
ID	ABR63268 standard; protein; 375 AA.						
XX	ABR63268;						
AC	AC						
CC	CC						
DT	28-AUG-2003 (first entry)						
XX	AD7c-NTP protein.						
DE							
XX	Cytostatic; Antibacterial; Immunosuppressive; Antiinflammatory;						
KW	neural thread protein; NTP; tumour.						
KW							
XX	Unidentified.						
OS							
XX							
FN	WO2003008443-A2.						
XX							
PD	30-JAN-2003.						


```
Db 69 GluLeuProThrSerAspProSerValSerAlaSerGlnSerAla-ArgTyrArg-T 88
QY 724 TGAGCCACCAGTCCCAA----- 708
Db 88 hrGlyHisHisAlaArgLeuCysLeuAlaAsnPheCysGlyArgAsnArgValSerLeuM 108
QY 707 -----CCTAGTTTCTCT---ACTTTAAAGCAT-----GTATTTTAAATAATAT 668
Db 108 etCysProSerTrpSerProGluLeuLysGlnSerThrCysLeuSerLeuProLysCysT 128
QY 667 GGAACATCCTTAATGAATACACTACATAGACATTTAGAAA----- 627
Db 128 rpAsp----- 139
QY 626 -----ATTAGGTATGCTGC---ACATGGAATTTTCCCAATAA--- 588
Db 139 helLeuPhePheLeuArgHisArgCysProThrLeuThrGlnAspGluValGlnTrpC 159
QY 587 -----AAAGCTAGTGTATGCCCAAAATCTGGAG-----G 557
Db 159 ysAspHisSerSerLeuGlnProSerThrProGluLeuLysHisProProAlaSerAla 179
QY 556 AATATGTAACCTGT-----TATCTTGGGTAGTAGGATTTATAGATGC 515
Db 179 erGlnValAlaGlyThrLysAspMetHisHisTyrThrTrpLeuIlePheIle 196
QY 514 CTTCACCTTCCCAATAGATATTTATGTAGAGTTTGAAGTTGCATTACT----- 465
Db 197 -----PheIlePheAsnPheLeuArgGlnSerLeuAsnSerValThrGlnAlaGlyV 214
QY 464 -----TTTGGCTTGTACT- 450
Db 214 alGlnTrpArgAsnLeuGlySerLeuGlnProLeuProGlyPheLysLeuPheSerC 234
QY 449 -----TTTA 446
Db 234 ysProSerLeuLeuSerSerTrpAspTyrArgArgProProArgLeuAlaAsnPhePheV 254
QY 445 TAACCAAAATAACAATAAGATTACAGTTAAATGCATTATAACTTTATTACA-----T 392
Db 254 alPheLeuValGluMetGlyPheThrMetPheAlaArgLeuLeuLeuSerGlyProC 274
QY 391 ATAATTACCTATCTCTGCTTACCTTCAGGAATCTTATTAGTTTGTCTCTCCGATCC 332
Db 274 ysAspLeuProAlaSerAlaSerGlnSerAlaGly-IleThrGlyValSerHisHisAla 293
QY 331 TTTATTTTATTATTAT---TTTTTGATAGAGTCTCGTGTGCGCCAGGCTGGA 275
Db 294 ArgLeuIlePheAsnPheCysLeuPheGluMetGluSerHisSerValThrGlnAlaGly 313
QY 274 GTCCAGTGTGGATCTCGCTCACTGCAACCTCGCTCCAGGTTTCAGCAATTTCT 215
Db 314 ValGlnTrpProLeuLeuGlySerLeuGlnProLeuProGlyLeuLysArgPheSer 333
QY 214 TGCTCAGCTCCCGATAGTGGACTACAGGATGCATCACTACACCCAGCTAATTT 155
Db 334 CysLeuSerLeuProSerSerTrpAspTyrGlyHisLeuProProHisProAlaAsnPhe 353
QY 154 TGTATTTTATAGATAGAGTTTACCATATGACAGGCTGGTCTCGAATTTCTCAC 95
Db 354 CysIlePheIleArgGlyGlyValSerProTyrLeuSerGlyTrpSerGlnThrProAsp 373
QY 94 CTCAAG 89
Db 374 LeuArg 375
RESULT 15
ID ABB99774
XX ABB99774 standard; protein; 375 AA.
AC ABB99774;
XX
```

```
DT 24-MAR-2003 (first entry)
XX Amino acid sequence of human neuronal thread protein AD7c-NTP.
DE Human; neuronal thread protein; AD7c-NTP; Alzheimer's disease; histone;
XX neurodegeneration; in vivo gene expression; amphipathic compound;
KW Homo sapiens.
OS WO200299036-A2.
PN 12-DEC-2002.
PD 28-MAY-2002; 2002WO-US016429.
PF 01-JUN-2001; 2001US-00872968.
PR (RHOD-) RHODE ISLAND HOSPITAL.
PA Wands JR, De La Monte SM;
XX WPI; 2003-140605/13.
XX N-PSDB; ABZ23236.
PS Inducing prolonged in vivo gene expression in mammal by contacting
XX neuronal tissue with composition comprising Alzheimer's disease-
XX associated neural thread protein 7c antisense nucleic acid, histone,
XX amphipathic compound.
XX Disclosure; Page 35; 69pp; English.
XX The present sequence represents a human neuronal thread protein AD7c-NTP.
XX AD7c-NTP is overexpressed in brains with Alzheimer's disease at early and
XX intermediate stages of neurodegeneration. The expression of AD7c-NTP may
XX be reduced using the method of the invention. The specification describes
XX a method for inducing prolonged in vivo gene expression in a mammal. The
XX method comprises contacting a non-muscular tissue with a composition
XX comprising a nucleic acid, histone and an amphipathic compound. The
XX method is useful for inducing prolonged in vivo gene expression in non-
XX muscular tissue of a mammal, e.g. neuronal tissue, central nervous system
XX (CNS) tissue, tissue comprising a post-mitotic neuronal cell, cortical
XX neuronal cell or hippocampal neuronal cell, glial cell, or vascular
XX endothelial cell. The method is useful in gene therapy applications to
XX treat Alzheimer's disease, where the composition comprises antisense
XX AD7c-NTP nucleic acid
XX Sequence 375 AA;
SQ
Alignment Scores:
Pred. No.: 5,58e-39 Length: 375
Score: 389.50 Matches: 128
Percent Similarity: 46.2% Conservative: 40
Best Local Similarity: 35.2% Mismatches: 88
Query Match: 23.8% Indels: 108
DB: 6 Gaps: 15
US-10-607-806-1-T9182_COPY_8700_9600 (1-901) x ABB99774 (1-375)
QY 901 GATACCTCTCGCTCAGACTCTCTAGTAGTGGAGTCTACAGGTCTTGCACCATGCCAG 842
Db 29 AspSerProAlaSerAlaSerProValAlaGlyIleThrGlyMetCysThrHisAlaArg 48
QY 841 CTAAT---TTTGTATTTTATAGAGGGTGTTCATCATGTTCAGGCTGTCTC 785
Db 49 LeuIleLeuTyrPhePheLeuValGluMetGluPheHisValGlyGlnAlaGlyLeu 68
QY 784 AAATCCCGGACCTCAAGTATTCATTCCTCGGCTCCCAAGTCTGGATACAGCG 725
Db 69 GluLeuProThrSerAspAspProSerValSerAlaSerGlnSerAla-ArgTyrArg-T 88
QY 724 TGAGCCACCAGTCCCAA----- 708
```

```
Db      88 hrGlyHisHisAlaAArgLeuCysLeuAlaAsnPheCysGlyArgAsnArgValSerLeuM 108
Qy      707 -----CCTAGTTTCTCT---ACTTTAAAGCAT-----GTATTTTAAAAATAT 668
Db      108 etCysProSerTrpSerProGluLeuLysGlnSerThrCysLeuSerLeuProLysCys 128
Qy      667 GGAACATCCITTAATGAATACTACATAGACATTTAGAAAA----- 627
Db      128 rAsp-----TyrArgAlaAlaValProGlyLeuP 139
Qy      626 -----ATTAGTATGCTGC---ACATGGAAATTTTCCCAATAAAA--- 588
Db      139 heIleLeuPheLeuArgHisArgCysProThrLeuThrGlnAspGluValGlnTrpC 159
Qy      587 -----AAAGCTAGTGTATGCCCAAAATCTGGAG-----G 557
Db      159 yAspHisSerSerLeuGlnProSerThrProGluIleLysHisProProAlaSerAla 179
Qy      556 AATATGTAACCTGGT-----TATCTTTGGGTAGTAGGATTATAGATGC 515
Db      179 erGlnValAlaGlyThrLysAspMethHisHisIleThrTrpLeuIlePheIle 196
Qy      514 CTTCACATTTCCATAATAGATATTTATGTAGAGTTTGAAGTTGCATTACT----- 465
Db      197 -----PheIlePheAsnPheLeuArgGlnSerLeuAsnSerValThrGlnAlaGlyV 214
Qy      464 -----TTTGGCTTGTACT- 450
Db      214 alGlnTrpArgAsnLeuGlySerLeuGlnProLeuProGlyPheLysLeuPheSerC 234
Qy      449 -----TTTA 446
Db      234 yProSerLeuLeuSerSerTrpAspTyrArgArgProProArgLeuAlaAsnPhePheV 254
Qy      445 TAACCAAAATAACAATAAGATTACAGTTAAATGCATTTAAACTTTATTACA-----T 392
Db      254 alPheLeuValGluMetGlyPheThrMetPheAlaArgLeuIleLeuIleSerGlyProC 274
Qy      391 ATAATTACGTATCTCTGCTTACCTTCAGGAATCTTATTAGTTTGTCTCTCCGATCC 332
Db      274 yAspLeuProAlaSerAlaSerGlnSerAlaGly-IleThrGlyValSerHisHisAla 293
Qy      331 TTTATTTTATTATTATTAT--TTTGTGATAGAGTCTGCTGTCGCCCGCTGGA 275
Db      294 ArgLeuIlePheAsnPheCysLeuPheGluMetGluSerHisSerValThrGlnAlaGly 313
Qy      274 GTGAGTGGTGGGATCTCGGCTCACTGCAACCTCTGCTCCAGGTTCAAGCAATTTCT 215
Db      314 ValGlnTrpProAsnLeuGlySerLeuGlnProLeuProGlyLeuLysArgPheSer 333
Qy      214 TGCCTCAGCTCCCGAGTAGTGGGACTACAGGATGCATCCTACACCCAGCTAATTTT 155
Db      334 CysLeuSerLeuProSerSerTrpAspTyrGlyHisLeuProProHisProAlaAsnPhe 353
Qy      154 TGTATTTTGTAGATAGAGTTTCCACATATTGACCAGGCTGCTCTCGAACTTCTCAC 95
Db      354 CysIlePheIleArgGlyGlyValSerProTyrLeuSerGlyTrpSerGlnTrpProAsp 373
Qy      94 CTCAAG 89
Db      374 LeuArg 375
```

Search completed: February 8, 2006, 10:42:44
Job time : 135.475 secs

GenCore version 5.1.7
Copyright (c) 1993 - 2006 Bioceleration Ltd.

OM nucleic - protein search, using frame_plus_n2p model

Run on: February 8, 2006, 10:44:07; Search time 6.14867 Seconds
(without alignments)
2819.838 Million cell updates/sec

Title: US-10-607-806-1-T9182_COPY_8700_9600

Perfect score: 1640

Sequence: 1 tggctcatattaataaaat.....gagctgagcaggagtatc 901

Scoring table:

BLOSUM62
Xgapop 10.0, Xgapext 0.5
Xgapop 10.0, Xgapext 0.5
Fgapop 6.0, Fgapext 7.0
Delop 6.0, Delext 7.0

Searched: 283416 seqs, 96216763 residues

Total number of hits satisfying chosen parameters: 566832

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Command line parameters:

-MODEL=frame+ n2p model -DEV=xlp
-Q/abs/ASSWEB_spool/MARTIN10607806/runat_08022006_092020_27698/app_query.fasta_1
-DB=PIR -QFMT=fastcan -SUFFIX=n2p.rpr -MINMATCH=0.1 -LOOPCL=0 -LOOPEXT=0
-UNITS=bits -START=1 -END=1 -MATRIX=blosum62 -TRANS=human40.cdi -LIST=45
-DOCALLIGN=200 -THR SCORE=pct -THR MAX=100 -THR MIN=0 -ALIGN=15 -MODE=LOCAL
-OUTFMT=ptc -NORM=ext -HEAPSIZE=500 -MINLEN=0 -MAXLEN=2000000000 -HOST=abs02p
-USER=MARTIN10607806 @CEN 1.1.92 @runat_08022006_092020_27698 -NCPU=6 -ICPU=3
-NO MMAP -NEG SCORES=0 -WAIT -SPELOCK=100 -LONGLOG -DEV_TIMEOUT=120
-WARN_TIMEOUT=30 -THREDS=1 -XGAPOP=10 -XGAPEXT=0.5 -FGAPOP=6 -FGAPEXT=7
-YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database :

1: PIR1:*
2: PIR2:*
3: PIR3:*
4: PIR4:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
C 1	388	23.7	613	4 C40201	artifact-warning s
C 2	357.5	21.8	673	4 F40201	artifact-warning s
C 3	352	21.5	627	4 A40201	artifact-warning s
C 4	350	21.3	673	4 F40201	artifact-warning s
C 5	335	20.4	627	4 A40201	artifact-warning s
C 6	327	19.9	613	4 C40201	artifact-warning s
C 7	318.5	19.4	418	2 S41044	chromosomal protei
C 8	286	17.4	597	4 E40201	artifact-warning s
C 9	264	16.1	196	2 I38022	hypothetical prote
C 10	256.5	15.6	597	4 E40201	artifact-warning s
C 11	232.5	14.2	100	2 A46010	X-linked retinopat
C 12	214.5	13.1	499	2 S65657	alpha-1C-adrenergi
C 13	182	11.1	579	4 D40201	artifact-warning s
C 14	173.5	10.6	579	4 D40201	artifact-warning s

C 15	172	10.5	46	2 I54375	gene NF2 protein -
C 16	171	10.4	301	4 B40201	artifact-warning s
C 17	166	10.1	447	2 A57034	transcription fact
C 18	162	9.9	39	2 I54374	gene NF2 protein -
C 19	160	9.8	369	2 A53959	thromboxane A-2 re
C 20	156	9.5	53	2 A42442	integrin beta-1 ch
C 21	155.5	9.5	407	2 T02670	probable thromboxa
C 22	154.5	9.4	79	2 A56194	thromboxane A-2 re
C 23	152	9.3	301	4 B40201	artifact-warning s
C 24	148.5	9.1	114	2 JC5238	galactosylceramide
C 25	147	9.0	841	1 I78885	serine/threonine-s
C 26	115.5	7.0	619	2 A60646	transforming prote
C 27	113.5	6.9	124	2 A47582	B-cell growth fact
C 28	109.5	6.7	440	2 A26359	decay-accelerating
C 29	108.5	6.6	331	2 A54295	interferon alpha/b
C 30	108.5	6.6	331	2 S59501	interferon recepto
C 31	101.5	6.2	107	2 C48078	biliary glycoprote
C 32	91	5.5	252	2 T48681	hypothetical prote
C 33	91	5.5	522	2 T08711	gamma-adaptin homo
C 34	90	5.5	339	2 S09264	Ig alpha chain C r
C 35	89	5.4	347	2 S09274	Ig alpha chain C r
C 36	83	5.1	352	2 S09266	Ig alpha chain C r
C 37	81.5	5.0	1559	1 S64757	probable membrane
C 38	80.5	4.9	350	2 T15178	hypothetical prote
C 39	80.5	4.9	1231	2 T24415	hypothetical prote
C 40	80	4.9	325	2 T26180	hypothetical prote
C 41	80	4.9	342	2 PC4211	hepatocellular car
C 42	79	4.8	101	2 S60434	probable membrane
C 43	79	4.8	348	2 S09270	Ig alpha chain C r
C 44	78	4.8	357	2 S09267	Ig alpha chain C r
C 45	78	4.8	1148	2 A71446	hypothetical prote

ALIGNMENTS

RESULT 1

C40201
artifact-warning sequence (translated ALU class C) - human

C;Species: Homo sapiens (man)

C;Date: 31-Mar-1992 #sequence_revision 11-Aug-1995 #text_change 19-May-2000

C;Accession: C40201

R;Claverie, J.M.

personal communication, 1992

A;Reference number: A40201

A;Accession: C40201

A;Molecule type: DNA

A;Residues: 1-613 <CLA>

A;Cross-references: UNIPARC:UPI000017CECA

R;Claverie, J.M.

Genomics 12, 838-841, 1992

A;Title: Identifying coding exons by similarity search: Alu-derived and other potential

A;Reference number: A40200; PMID:92241891; PMID:1572661

C;Comment: This "warnings" entry is a conceptual translation in all 6 reading frames of

in-frame stop codons are shown as 'X'.

C;Comment: Any significant similarity of a predicted protein sequence to a portion of t

Alignment Scores:

Pred. No.:	2.6e-35	Length:	613
Score:	388.00	Matches:	119
Percent Similarity:	50.8%	Conservative:	36
Best Local Similarity:	39.0%	Mismatches:	100
Query Match:	39.7%	Indels:	51
DB:	4	Gaps:	8

US-10-607-806-1-T9182_COPY_8700_9600 (1-901) x C40201 (1-613)

QY	899	TACTCTGCTCAGACTCTGAGTACCTGGACTCAGGTGCTTGCACATGCCAGCT	840
Db	345	PhSeSerCysLeuSerLeuLeuArgThrTrpAspTyrArg-ArgProHisalaGlnLe	364
QY	839	AAT-TTTTGTATTTTATTATTAGAGAGGGTGTTCATCATGTGTCCAGGCTGCTCAAAC	781

Db 596 rPheGlySerLeuGlnProGlnProGlyPhePhyArgPheSerCysLeuCysLeuPr 616
Qy 200 GAGTAGCTGGGACTACAGCATGCATCACTACCCAGCTAATTTTGTATTTTAGTAG 141
Db 616 oCysSerTrpAspTrpArgHisGlyProProPheLeuAlaAsnLeuCysIlePheAsnAr 636
Qy 140 AGATAGGTTTCCACCATATTGACCAGCGTGTCTCGAACTTCTCACCTCAAGTGCAC 81
Db 636 gAspThrValSerProCys***SerGlyTrpSerGlnThrProAspLeuLys***SerAl 656
Qy 80 CCACCTTGGACTTCCAAAGTACTGTGTGTACAGGCTTGAGTCACTGTCGCCAA 28
Db 656 a-LeuLeuValSerGlnSerAlaGlyMetThrGlyValSerHisHisThrGln 673
RESULT 3
A40201
artifact-warning sequence (translated ALU class A) - human
C;Species: Homo sapiens (man)
C;Date: 31-Mar-1992 #sequence_revision 11-Aug-1995 #text_change 19-May-2000
C;Accession: A40201
R;Claverie, J.M.
personal communication, 1992
A;Reference number: A40201
A;Accession: A40201
A;Molecule type: DNA
A;Residues: 1-627 <CLA>
A;Cross-references: UNIPARC:UPI000017CEC8
R;Claverie, J.M.
Genomics 12, 838-841, 1992
A;Title: Identifying coding exons by similarity search: Alu-derived and other potential
A;Reference number: A40200; MUID:92241891; PMID:1572661
A;Contents: annotation
C;Comment: This "warning" entry is a conceptual translation in all 6 reading frames of c
in-frame stop codons are shown as 'x'.
C;Comment: Any significant similarity of a predicted protein sequence to a portion of th

Alignment Scores:
Pred. No.: 3.2e-31 Length: 627
Score: 352.00 Matches: 117
Percent Similarity: 50.2% Conservative: 36
Best Local Similarity: 38.4% Mismatches: 99
Query Match: 21.5% Indels: 53
DB: 4 Gaps: 11

US-10-607-806-1-T9182_COPY_8700_9600 (1-901) x A40201 (1-627)

Qy 899 TATCTGCTCAGACTCCTGAGTAGTGGAGTACAGGTGCTTGCCACCATGCCAGCT 840
Db 355 PheSerCysLeuSerLeuProSerSerTrpAspTrpArgProHisTyAlaArgLeu 374
Qy 839 AATTTTGTATTTTATTAGAGGGGTGTTTCATCATGTTGTCAGGCTGCTCAAACT 780
Db 375 IlePheCysIlePheSerArgAspGlyValSerProTrp***ProGlyTrpSerArgSer 394
Qy 779 CCGACCTCAAGTATCCATTTGCTCGGCTCCCAAGTGGGATACAGCGGTGAGC 720
Db 395 ProAspLeu--ValIleArgProProTrpProProLysValLeuGlyLeuGlnAla***A 414
Qy 719 CACATGCCCAACTAGTTCTCTACTTTAAAGCATGATTTTAAATAATTTGGAACATC 660
Db 414 spArgAlaArg*****PhePheIlePheCys----- 425
Qy 659 CTTAATGGAATACATAGATAGATTAGAAAATAGGTATGCTGCATCATGGAATTT 600
Db 426 -----PheTy----- 429
Qy 599 TCCCAATAAATAAAGCTAGTGTATGCC-----CAAAAATCTGGAGGAATA----- 553
Db 429 heProArgArgSerLeuAlaLeuSerProArgLeuGluCysSerGlyAlaIleSerAlah 449
Qy 552 --TGTAACTGGTTATCTTTGGGTAGTAGGATTATAGATGCTTCCACTTTCCATAATAGA 495
Db 552 ----- 495

Db 449 iScysLysLeuArgLeuProGlySerArgHisSerProAla-SerAla-----SerGln 466
Qy 494 TATTATTAGTAGATTGGAAGTTGCATTTACCTTTTGGC---TTGTTTACTTTTATTAACCAA 438
Db 467 ValAlaGlyThrThrGlyAlaArgThrThrProGly***PhePheValPheLeuValGlu 486
Qy 437 ATAACAATAAGATTACAGTTAAATGCATTTAAACCTTTTATTACA---TATAATTTACGT 381
Db 487 ThrGlyPheHisArgGlySerGlnAspGlyLeuAspLeuLeuThrSer***SerAlaArg 506
Qy 380 ATCTCTGCT-----TACCTTTCAGAAATCTTTATTATTAGTTTGTCTCTCCG 336
Db 507 LeuGlyLeuProLysCysTrpAspTyArgArgGluThr-----AlaPro 521
Qy 335 -----ATCCTTTATTATTATTATTATTATTATTATTGAGATAGAGTCTCGCTGTGT 288
Db 522 Gly*****PheLeuPhePheValPheIlePhePheArgAspGlyValSerLeuCys 541
Qy 287 CGCCAGCGCTGGAGTGCAGTGGTGGGATCTCGGCTCACTGCAACCTCTGCTTCCAGGTT 228
Db 542 ArgGlnGlyTrpSerAlaValAlaArgSerArgLeuThrAlaSerSerAlaSerArgVal 561
Qy 227 CAAGCAATCTCTGCTCAGCTCCGAGTACGAGTACGAGTACGAGCATCATCACTACA 168
Db 562 HisAlaIleLeuLeuProGlnProProLys***LeuGlyLeuGlnAlaProAlaLeuArg 581
Qy 167 CCCAGCTAATTTTGTATTTTATTAGTAGATAGAGTGGTTTACCATATTTGACCAGGCTGTC 108
Db 582 ProAlaAsnPheLeuTyPhe*****ArgArgGlyPheThrValValAlaArgMetVal 601
Qy 107 TCCAACCTTCTCACTCAAGTGCACCCACCTTGGACTTCCAAAGTACTGTGTTGTACAG 48
Db 602 SerIleSer***ProArg--AspProProAlaLeuAlaSerGlnSerAlaGlyIleThrG 621
Qy 47 GCTTG 43
Db 621 lyVal 622

RESULT 4
F40201
artifact-warning sequence (translated ALU class F) - human
C;Species: Homo sapiens (man)
C;Date: 31-Mar-1992 #sequence_revision 11-Aug-1995 #text_change 19-May-2000
C;Accession: F40201
R;Claverie, J.M.
personal communication, 1992
A;Reference number: A40201
A;Accession: F40201
A;Molecule type: DNA
A;Residues: 1-673 <CLA>
A;Cross-references: UNIPARC:UPI000017CECD
R;Claverie, J.M.
Genomics 12, 838-841, 1992
A;Title: Identifying coding exons by similarity search: Alu-derived and other potential
A;Reference number: A40200; MUID:92241891; PMID:1572661
A;Contents: annotation
C;Comment: This "warning" entry is a conceptual translation in all 6 reading frames of
in-frame stop codons are shown as 'x'.
C;Comment: Any significant similarity of a predicted protein sequence to a portion of t

Alignment Scores:
Pred. No.: 5.43e-31 Length: 673
Score: 350.00 Matches: 116
Percent Similarity: 47.0% Conservative: 33
Best Local Similarity: 36.6% Mismatches: 114
Query Match: 21.3% Indels: 54
DB: 4 Gaps: 6

US-10-607-806-1-T9182_COPY_8700_9600 (1-901) x F40201 (1-673)

Qy 29 TGGGCACAGTACTCAAGCTGTGAACCAACAGTACTTTGGAAGTCCAAAGGTGGGTGATCA 88
Db 29 TGGGCACAGTACTCAAGCTGTGAACCAACAGTACTTTGGAAGTCCAAAGGTGGGTGATCA 88

```

Db      2 TrpValTrpTrpLeuThrProValIleProAlaLeuTrpGluThr-ArgArgAlaAspHi 21
Qy      89 CTGAGGTGAGAGTTCGAGACGAGCTGTGTCATATGTCGAACCCCTATCTCTACTAAA 148
Db      21 sLeuArgSerGlyVal***AspGlnProAspGlnHisGlyGluThrValSerLeuLeuLy 41
Qy      149 AATACAAATAATAGCTGGGTAGTATGATGATGCTGCTAGTCCAGCTACTCCGGAGGCT 208
Db      41 sIleHisLeuAlaAlaArgGlnGlyProCysLeu***SerGlnLeuHisGlyArgG1 61
Qy      209 GAGCAAGAGAGAAATGCTTTGAACCTGGGAGGAGAGTTGCAGTCGAGCGAGATCCCAACCA 268
Db      61 nArgGlnGluAsnArgLeuAsnProGlyGly***GlyCysSerGluProLysLeu-Alat 81
Qy      269 CTGACCTCCAGCTTCGGGGACACAGCAGAGACTCTATCTCAAAAAATAAATAAATAAAT 328
Db      81 hrAlaLeuGlnProGlyCysGlnSerLysGlyLeuSerGlnLysGlnLysGlnSerLys- 100
Qy      329 AAAGGATCGGAGACAAACAACTAATAAG----- 358
Db      101 LysLysLysLysLysThrProLysAsnLys*****AlaGlyCysGlyGlySerArg 120
Qy      359 -----ATTCCTGAAGTAAGCAGAGATAC 382
Db      121 LeuSerSerGlnHisPheGlyArgProGlyGlyGlnIleThr***GlyGlnGluPheGlu 140
Qy      383 GTAAATATATGTAATAAAGTTTAAATGCATTTTAACTGTAATCTTATTTATTTTGG 442
Db      141 ThrSer-LeuIleAsnMetVal-----LysLeuCysLeuTy***Ly 154
Qy      443 TTATAAAGTAACAACGCAAAAGTAATGCAACTTCAAACTCTACATAAATATCTATTAT 502
Db      154 sTyIleAsn***ProGlyMetAlaAlaHisAlaCysAsnProSerTyThrGly----- 172
Qy      503 GGAAGTGGAAAGGCATCTA-----TAATCTACTACCCAAAGATAACCACTTACAT 553
Db      173 -AspArgGlyArgArgIleAla***ThrGlnGluValGluAlaValSerGlnAsnLe 192
Qy      554 ATTCTTCAGATTTTGGGGCATACATAGCTGTTTTTTTATTTGGGAAATTTCCATGTGC 613
Db      192 uProLeuHisSerSerArgGlyValArgAlaArgAlaTyLeuLysAsn----- 208
Qy      614 AGGCATACCTAATTTTCTAAATGCTATGATGATGATTCATTTAAGGATGTTCCATAAT 673
Db      209 -----LysAsnLysAlaLysLysLysLysLysProLys----- 221
Qy      674 TTTAAATACATGCTTTTAAAGTAGAGAACTAGGTGGGCATGTGGCTCACGCTGTAT 733
Db      222 -----ThrLys*****LeuGlyValValAlaHisAlaCysHi 235
Qy      734 CCCAGCACTTTGGAGCGCGAGGCAATGGATGATCATTGAGTCCGGAGTTTGACACGAC 793
Db      235 sProSerThrLeuGlyAspGlnGluGlyArgSerLeuGluValArgSerLeuArgProAl 255
Qy      794 CTGACACATGATGAAACCCCTCTCTAAATAAAATAACAAATATAGCTGGCATGTG 853
Db      255 a***SerThrTrp***AsnCysValSerIleLysAsnThr***IleSerGlnGluTrpAr 275
Qy      854 GCAAGCACTGTAGTCCAGCTACTCAGGAGCTCTGAGGCAGGA 896
Db      275 gProMetProValIleProAlaThrArgGluThrGluAlaGly 289

```

RESULT 5
A40201
artifact-warning sequence (translated ALU class A) - human
C:Species: Homo sapiens (man)
C:Date: 31-Mar-1992 #sequence_revision 11-Aug-1995 #text_change 19-May-2000
C:Accession: A40201
R:Claverie, J.M.
personal communication, 1992
A:Reference number: A40201
A:Accession: A40201
A:Molecule type: DNA

A:Residues: 1-627 <CLA>
A:Cross-references: UNIPARC:UPI000017CEC8
R:Claverie, J.M.
Genomics 12, 838-841, 1992
A:Title: Identifying coding exons by similarity search: Alu-derived and other potentially
A:Reference number: A40200; PMID:92241891; PMID:1572661
A:Contents: annotation
C:Comment: This "warning" entry is a conceptual translation in all 6 reading frames of or
in-frame stop codons are shown as 'X'.
C:Comment: Any significant similarity of a predicted protein sequence to a portion of the

Alignment Scores:	2,73e-29	Length:	627
Pred. No.:	335.00	Matches:	116
Score:	43.7%	Conservative:	27
Best Local Similarity:	35.5%	Mismatches:	87
Query Match:	20.4%	Indels:	97
DB:	4	Gaps:	8

US-10-607-806-1-T9182_COPY_8700_9600 (1-901) x A40201 (1-627)

```

Qy      40 ACTCAAGCCTGTAAACACAGACTACTTTGGAAGTCAAGGTGGGTGGATCACTTGAGGTGAG 99
Db      5 SerHisAlaCysAsnProSerThrLeuGlyGlyGlnGlyArgIleThr--ArgSerG 24
Qy      100 AAGTTTCGAGACAGCCTGCTCAATATGTTGAAACCCCTATCTCTACTAAATAACAAAT 159
Db      24 lyAspArgAspHisProGlyTyThrHisGlyGluThrProSerLeuLeuLysIleGlnLysI 44
Qy      160 TAGCTGGGTGTAGTGTGATGCTGCTAGTCCAGCTACTCGGAGGCTCGGCAAGCAGA 219
Db      44 leSerArgAla***CysGlyArgLeu***SerGlnLeuLeuGlyArgLeuArgGlnGluA 64
Qy      220 ATTGCTTGACCTGGAGGAGGAGTGTGAGTGGAGGAGATCCACCATCGACTCCAG 279
Db      64 snGlyValAsnProGlyGlyGlyAlaCysSerGluProArgSerArgHisCysThrProA 84
Qy      280 CTGGGGGACACAGCAGACTCTATCTCAAAAAATAAATAAATAAATAAAGATCGGA 339
Db      84 laLeuAlaThrGluArgAspSerValSerGlnLys-----AsnLysAsnLysLys***L 102
Qy      340 GAGAAACAAACTAATAAGATTCCTGAAAGTAAAGGTAAGCAGAGATACGTAAATTTATGTAATA 399
Db      102 ys*****AlaGlyArgGlyLeu----- 110
Qy      400 AAGTTTAAATGCATTTTAACTGTAATCTTATGTTTATTTGTTTATATAAAGTAAACAAG 459
Db      111 -----ThrProValIleProAlaLeuTrpGluAlaLysAlaG 123
Qy      460 CCAAAAGTAATGCNACTTCAAACTCTACATAAATAATCTATTTATGGAAGTGAAGGCATC 519
Db      123 lyGlySerArgGlyGlnGluIleGluThrIleLeuAlaThrValLysProArg--L 142
Qy      520 TATAATCCTACTCCCAAGATAACCAAGTATACATATTTCTCCAGATTTTGGGCATACA 579
Db      142 euTy***LysTyLysLysLeuAla----- 150
Qy      580 CTAGCTTTTATTTATTTGGGAAATTTCCATGTGCGAGCATACCTAATTTTCTAAATGTC 639
Db      151 -----GlyArg-----SerAlaGlyA 156
Qy      640 TATGTAGTATTCATTTAAGGATGTTCCATAATTTTFA-----AAATACATGC 687
Db      156 laCysSer-----ProSerTyLeuGlyGly***GlyArgArgMeta 170
Qy      688 TTTAAAGTAGAGAACTAGGTG----- 710
Db      170 la***ThrArgGluAlaGluLeuAlaValSerArgAspAlaThrAlaLeuGlnProT 190
Qy      710 ----- 710
Db      190 rpArgGlnSerGluThrProSerArgLysLysIleLysThrLysAsnLysLys***** 210

```


Qy	711	-----GGCATGGTGCTCACGCCTCATCCAGCACTTTGGAGGCGCAGGCAAAATGGAT	765
Db	210	***ProGlyAlaValSerArgLeu-***SerGlnHisPheGlyArgProArgArgAlaAsp	229
Qy	766	CACTTTGAGTCCGGAGTTTGAGACCAAGCTGCAACAACATGATGAACACCCCTCTCTAATA	825
Db	230	His--GluValArgSerArgProSerTrpLeuProArg***AsnProValSerThrL	249
Qy	826	AAAAATACAAAAATTAGCTGGGCATGGTGGCAAGCACCTGTAGTCCAGGTACTCAGGAGT	885
Db	249	ysAsnThrLysAsn***ProGlyValValArgAlaProValValProAlaThrTrpGluA	269
Qy	886	CTGAGGCAGGA	896
Db	269	laGluAlaGly	272

RESULT 6
C40201
artifact-warning sequence (translated ALU class C) - human
C:Species: Homo sapiens (man)
C:Date: 31-Mar-1992 #sequence_revision 11-Aug-1995 #text_change 19-May-2000
C:Accession: C40201
R:Claverie, J.M.
personal communication, 1992
A:Reference number: A40201
A:Accession: C40201
A:Molecule type: DNA
A:Residues: 1-613 <CLA>
A:Cross-references: UNIPARC:UPI000017CECA
R:Claverie, J.M.
Genomics 12, 838-841, 1992
A:Title: Identifying coding exons by similarity search: Alu-derived and other potentially
A:Reference number: A40200; MUID:92241891; PMID:1572661
C:Contents: annotation
C:Comment: This "warning" entry is a conceptual translation in all 6 reading frames of c
in-frame stop codons are shown as 'X'.
C:Comment: Any significant similarity of a predicted protein sequence to a portion of the

Alignment Scores:	
Pred. No.:	2.2e-28
Score:	327.00
Length:	613
Percent Similarity:	116
Matches:	34
Conservative:	101
Best Local Similarity:	37.8%
Mismatches:	57
Query Match:	19.9%
Indels:	9
DB:	4
Gaps:	9

US-10-607-806-1-T9182 COPY 8700 9600 (1-901) x C40201 (1-613)

QY 35 CAGTGACTCAAGCCTGTAAACCACAGTACTTTGGAAGTCACAGGTGGGTGGATCATCTGAG 94
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Dd 4 GlnTrpLeuThrProValIleProLaLeuTrpGluAlaIysValGlyGSer***--G 23

QY 95 GTGAGAAAGTTCGAGACCAGCCTGGTCAATATGGTGAAACCCCTATCTCTACTAAAAATACA 154

Db
23 1yGlnGluPheGluThrSerLeuAlaAsnMetAlaLysProHisPheTyr**LysTyrL 43

QY 155 AAAATTAGCTGGGTAGTGATGCATGCCCTGTAGTCCAGCTACTCGGGAGGCTGAGGCA 214

Db

43 ySLysLeuAlaGlyHisGlyAlaGlyAlaCvssSerProThrTyrAlaGluGly***GlyVA 63

QY 215 AGAGAAATTGCTTGAACCTGGAGGCCAGAGGTTGCAGTGAGCCGAGATCCACCACCTGCAC 274

Db
63 rgArgMetAla***ThrArgGluValGluLeuAlaValSerArgAspLeuAlaIleAlaL 83

QY 275 TCCAGCCTGGCGACACAGCGAGACTCTATCTCAAAAAATAAAATAAAAT----- 328

Db

83 euGlnProGly**GlnSerGluThrLeuSerGlnLysLys-LysLysLys***** 102

QY 329 ---AAAGGATCGGAGAGAAACAAAACCTAATAAGATTCTGAAGGTAAGCAGAGATACGTA 385

Db
103 ***AlaGlyCysSerGlySerArgLeu***SerGlnHisPheGlyArgGlnArgTrpVal 122

386	QY	AATTATATGTAATAAGATTAAATGTCATTTTAACTGTAATCTTATTTGTTATTTTGGTGA	445
123	Db	AspHisGluAlaArgSerSerArgPro-----AlaTrpPro	134
446	QY	TAAAGATAAACAAAGCCAAAGTAATGCAACTTCAAACTCTACATAAATATCTATTATGGA	505
135	Db	-----ThrTrpGlnAsnProIleSerThrLysAsnThrLysAsn-----LeuGly	150
506	QY	AAGTGAAGGCATCTATATCTACTACCCCAAAGATAACCACTTACATATT-----	556
151	Db	MetValArgAlaProValVal-----ProArgThrGlnLysAlaGluAlaGlyGluTrp	168
557	QY	---CCTCCAGATTTTGGGCG-----ATACACTAGCTTTTT	589
169	Db	HisLysProGlyArgTrpSerLeuGln**AlaGluIleSerProLeuHis--SerSer	187
590	QY	TTATTGGGAAAAATTCATGTGCAGGCATACCTTAATTTTCTAAATGCTCTATGTAGTAT	649
188	Db	LeuGlyAspArgValArgLeuCys-----	195
650	QY	TCCATTTAAGGATGTTCCATAATTTTAAATACATGCTTTAAAGTAGAGAAACTAGGTT	709
196	Db	-----LeuLysLysLysLysLysLys-----Lys*****Ar	207
710	QY	GGGCATGTTGCTCACGCTGTAT-CCGACGACTTTGGGAGCGCGAGGCAATGATGATCAC	768
207	Db	gGlyAlaValAlaHisAlaCysAsnProSerThrLeuGlyGlyLysGlyGlyTrpIle--	226
769	QY	TTGAGTCCGAGTTTGAGACCGCTGGACAAATGATGAACACCCCTCTCTTAATAAAA	828
227	Db	MetArgProGlyValArgAspGlnProGlyGlnHisGlyLysThrProPheLeuLeuLys	246
829	QY	ATACAAAAATTAGCTGGCGCATGTTGGCGAAGCACTCTAGTCCAGCTACTCAGGAGTCTG	888
247	Db	IleGlnLysIleSerTrpAlaTrpCysGlyArgLeu**SerHisValArgArgLeu	266
889	QY	AGGCAGGAG	897
267	Db	ArgGlnGlu	269

RESULT 7

S41044
chromo
C;Spec
C;Date
C;Acce

R; Yeo,

Nature

A; Refe:

A; Statv

A;Resio

A; C: Crows;
C: Sings.

C; Keywords

Alignme
Pred

Score:

Best Lo

DB:

US-10-0

QY

Db

•

[illegible]

RESULT 8

E40201
artfact-warning sequence (translated ALU class E) - human
C:Species: Homo sapiens (man)
C:Date: 31-Mar-1992 #sequence_revision 11-Aug-1995 #text_c
C:Accession: E40201
R:Claverie, J.M.
Personal Communication, 1992
A:Reference number: A40201
A:Accession: E40201
A:Molecule type: DNA
A:Residues: 1-597 <CLA>
A:Cross-references: UNIPARC:UPI000017CECC

R:Claverie, J.M.
Genomics 12, 838-841, 1992
A:Title: Identifying coding exons by similarity search: Alu-derived and other potentially
A:Reference number: A40200; MUID:92241891; PMID:1572661
A:Contents: annotation
C:Comment: This "warning" entry is a conceptual translation in all 6 reading frames of or
in-frame stop codons are shown as 'X'.
C:Comment: Any significant similarity of a predicted protein sequence to a portion of the

Alignment Scores:	9.98e-24	Length:	597
Pred. No.:	286.00	Matches:	116
Score:	45.0%	Conservative:	25
Percent Similarity:	37.1%	Mismatches:	74
Best Local Similarity:	17.4%	Indels:	102
Query Match:	4	Gaps:	10
DB:			
US-10-607-806-1-T9182_COPY_8700_9600 (1-901) x E40201 (1-597)			
Qy 898	ACTCTCGCTCCAGACTCTCAGTACTGCGGACTACAGGTCTTGTGCCACCATGCCAGCTA	839	
Db 334	SerProValSerAlaSerArgValAlaGlyIleThrGlyMetArgHisAlaGlnLeu	353	
Qy 838	ATTTTGTATTTTATTTAGAGAGGGTGTTCATCATGTTCGACGGCTGTGTCTCAAACTC	779	
Db 354	Ile-LeuTyrPhe*****ArgTrpGlyPheSerMetLeuValLeuValSerAsnSe	373	
Qy 778	CGGACCTCAAGTATCATTTT-GCCTCGGCTCCCAAAGTCTGGGAT-ACAGCGCTGAG	721	
Db 373	rGlnProGlnValIleArgProAlaLeuAsnSerGlnSerAlaGlyIleThrGlyMetSe	393	
Qy 720	CCACCATGCCAACCTAGTTTCTCTACTTTAAAGCATGTATTTTAAAAATTATGGAAACAT	661	
Db 393	rTyRHs-----	395	
Qy 660	CCTTAAATGGAATACTACATAGACATTTAGAAAAATTAGTATGCTTCGCACATGGAAA-	603	
Db 396	-----ThrTrp*****	399	
Qy 602	-----TTTTCCTCAATAAAAAAGCTAGTGTATGCCCCCAAAAATCTGGAGGAATA	553	
Db 399	***PhePhePheGluThrGluPheCysSerCysCysProGlyArg-----	415	
Qy 552	TGTAACGTGTTATCTTTGGGTAGTAGGATTATAGATGCCCTCCACCTTTC-----CATAA	499	
Db 416	-----SerThrMetAlaGlnSerHi	422	
Qy 498	TAGATATTTATGTAGACTTTTCAAGT-----	474	
Db 422	sArgLysLeuCysPheLeuGlySerSerAsnSerLeuLeuSerGlnProGlu***Le	442	
Qy 473	-----TGCATTACT-----TTTGGCTGTTTACTTTTATAACCA	439	
Db 442	uGlyLeuGlnAlaCysAlaThrThrProSer***PheTyrIlePheSer-----	458	
Qy 438	AATAAACAATAAGATACAGTTAAATGTCATTTAAACTTTTATTATCAATAATTTACGTAT	379	
Db 459	----ArgAspGlyValSerProCysTrpSerSerTrpSerGlnThrProAsnLeuArg**	477	
Qy 378	CTCTGCTTA-----CCTTCAGGAATCTTATTAGTTTGTGTTCTCTCCGATCCCTTT	329	
Db 477	*SerAlaPro***ThrProLysAlaLeuGlyLeuGlnAla***AlaThrThrProGln	497	
Qy 328	ATTTTATTTATTTTATTTTTCAGATAGACTCTCGCTGT-GTCCGCCAGCTCGAGTGCA	270	
Db 497	y*****PhePhePheLeuArgArgSerPheAlaLeuValAlaGlnAlaGlyValar	517	
Qy 269	GTGTGGGTATCTCGGCTCA-----CTGCAACCTCTGCTCCAGGTTCAA	225	
Db 517	gtrPHiASnLeuThrAlaAsnPheAlaSerTrpValGlnAlaIle-----	532	
Qy 224	GCAATTCCTTCGCTCCAGCTCCCGAGTAGCTGGGACTACAGGCATGCATCCTACACCC	165	

Db 533 ----LeuSerCysLeuSerLeuProSerSerTrpAspTyrArgHisAlaProProArgPr 551
164 AGCTAAATTTGTTTCTAGATAGAGTAGGTTTCCACATTTGACCGAGCTGCTCG 105
Db 551 oAlaAsn-PhelIeHeuValGluMetGlyPheLeuHisValGlyGlnAlaGlyLeu 571
Qy 104 AACTTCTCACCTCAAGTGATCCACCCACCTTGGAC 70
Db 571 ysLeuProThrSerGlyAspProProArgLeuGlu 582

RESULT 9
I38022
hypothetical protein - human
C/Species: Homo sapiens (man)
C/Date: 01-Nov-1996 #sequence_revision 01-Nov-1996 #text_change 05-Nov-1999
C/Accession: I38022
R/Yang, S.S.; Zhang, K.; Vieira, W.; Taub, J.V.; Zeilstra-Ryalls, J.H.; Somerville, R.L.
Cancer Res. 50, 5658-5667, 1990
A/Title: A human hepatocellular carcinoma 3.0-kilobase DNA sequence transforms both rat
A/Reference number: I38021
A/Accession: I38022
A/Status: preliminary; translated from GB/EMBL/DBJ
A/Molecule type: DNA
A/Residues: 1-196 <RES>
A/Cross-references: UNIPARC:UPI000011E1FB; EMBL:X55777; NID:G288143; PIDN:CAA39297.1; PI

Alignment Scores:
Pred. No.: 2,85e-21 Length: 196
Score: 264.00 Matches: 61
Percent Similarity: 68.2% Conservative: 12
Best Local Similarity: 57.0% Mismatches: 23
Query Match: 16.1% Indels: 12
DB: 2 Gaps: 3

US-10-607-806-1-T9182_COPY_8700_9600 (1-901) x I38022 (1-196)
Qy 87 CACTTGAGGTGAGAAAGTTCGAGACGACCTGGTCAATATGGTGAACCCCTCTACTA 146
Db 3 HisLeuArgSerGlyValGlnAspTyrProGlyGlnHisGlyLysIleProSerLeuLeu 22
Qy 147 AAAATCAAAAATAGCTGGGTGAGTGTGATGTCATGCTGTAGTCCCGAGTACTCGGAGG 206
Db 23 LysIleGlnGluLeuAlaGlyHisGlyGlyArgCysLeuGlnSerGlnLeuLeuArg 42
Qy 207 CTGAGGCAAGAAATCTCTGAACCTGGAGGAGAGGTTGCAGTGCAGTGCAGTGCAGTCCAC 266
Db 43 LeuArgGlnGluAsnHisLeuAsnSerGlyGlyArgCysSerGluProLysSerHis 62
Qy 267 CACTGCACCTCAGCTCGGCGACACGAGAGCTCTATCTCAAAAATAATAATAATAA 326
Db 63 LeuCysIleProAlaTrpValThrGluGlyAspSerValSerLysGln-----AsnLys 80
Qy 327 ATAAGATCGGAG-----AGAAACAAACTAATAAGATT----- 361
Db 81 ThrLysAsnGlnGlnHisLeuArgAsnAsnThrLysLysSerAsnSerCysIleIleGly 100
Qy 362 ---CCTGAAGGTAAACAGAGA 379
Db 101 GlyProGluGlyGluGluLys 107

RESULT 10
E40201
artifact-warning sequence (translated ALU class E) - human
C/Species: Homo sapiens (man)
C/Date: 31-Mar-1992 #sequence_revision 11-Aug-1995 #text_change 19-May-2000
C/Accession: E40201
R/Claverie, J.M.
personal communication, 1992
A/Reference number: A40201
A/Accession: E40201
A/Molecule type: DNA
A/Residues: 1-597 <CLA>

A/Cross-references: UNIPARC:UPI000017CECC
R/Claverie, J.M.
Genomics 12, 838-841, 1992
A/Title: Identifying coding exons by similarity search: Alu-derived and other potential
A/Reference number: A40200; MUID:92241891; PMID:1572661
A/Comments: annotation
C/Comment: This "warning" entry is a conceptual translation in all 6 reading frames of
in-frame stop codons are shown as 'X'.
C/Comment: Any significant similarity of a predicted protein sequence to a portion of t

Alignment Scores:
Pred. No.: 2,24e-20 Length: 597
Score: 256.50 Matches: 112
Percent Similarity: 45.5% Conservative: 26
Best Local Similarity: 37.0% Mismatches: 99
Query Match: 15.6% Indels: 67
DB: 4 Gaps: 10

US-10-607-806-1-T9182_COPY_8700_9600 (1-901) x E40201 (1-597)
Qy 41 CTCAGGCTGTAAACCACAGTACTTTGGAAGTCAAGTGGGTGG-ATCATTGAGGTGAG 99
Db 6 LeuMetProValIleProAlaLeuTrpGluPheLysAlaGlyArgIleThr***GlyTrp 25
Qy 100 AAGTTCGAGACCAAGCTGGTCAATATGTTGTAACCCCTCTCTCTACTAAAAATACAAAAAT 159
Db 26 GluPheGluThrSerLeuThrAsnMetGluLysProHisLeuTyr***LysTyr-LysIle 45
Qy 160 TAGCTGGGTGTAGTATGATGCTGTAGTCCAGAGTACTCTCGGAGGCTCAG---GCAAG 216
Db 45 eSerTrpAlaTrpTrpArgMetProValIleProAlaThrArgGluAlaGluThrGly-G 65
Qy 217 AGAATTGCTTGAACCTGGGAGGAGGAGTGTGAGTGCAGCGAGATCCACCATCGCACTC 276
Db 65 luAsnCysLeuAsnProGlySerLysValCysGlyGlu-IleValProSerTyrSer--- 83
Qy 277 CAGCCTGGGCGACACGAGAGTCTCTCTCAAAAATAATAATAATAATAAAGATC 336
Db 84 GlyLeuGlyAsnLysSerLysThrProSerGlnLysLysLys*****AlaArGcys 103
Qy 337 GGAGAGAAACAAACATAAAGATTCTCTGAAGGTAAGCAGAGATACGTAATATATGTA 396
Db 104 GlySer----- 105
Qy 397 ATAAAGTTAAATGCAATTTAACTGTAATCTTATTTGTTTATTTGTTTATAAAGTAAAC 456
Db 106 -----SerCysLeu***Ser-----GlnArgPheGlySerSerArgArgGlyGly 120
Qy 457 AAGCCAAAAGTAATGCAACTTCAAACTCTACAACTACATAATATCTATTATGGAAGTGAAGGC 516
Db 121 SerProGluValGlySerLeuArgProAla***Pro-----ThrTrpArg----- 135
Qy 517 ATCTATAATCCTACTACCCAA-----AGATAACCCAGTTACATAT 555
Db 136 -----AsnProIleSerThrLysAsnIleLysLeuAlaGlyArgGlyGlyAlaCysLeu 153
Qy 556 TCCTCCAGATTTTGGGGCATACACTAGCTTTTATTTTGGGAAAATTTCCATGTCGAC 615
Db 154 ***SerGlnLeuLeuGly-----ArgLeuArg 162
Qy 616 GCATACCTAATTTTCTAAATGCTATGTAGTATTTCATTTAAGGATGTTCCATATTTT 675
Db 163 GlnGluArgIleAla***ThrGlnGluAlaLysPheAlaValArgLeuCysHisArg-Th 182
Qy 676 TAAATATACATGCTTTTAAAGTAGAGAACTAGTGTG-----G 711
Db 182 rProAlaTrpAlaThrArg-AlaLysLeuArgLeuLysLysLysLys*****ProG 202
Qy 712 GCATGTGGCTCAGCGCTGTATCCCA---GCATTTGGGAGGCCGCGGCAATATGGATCAC 768
Db 202 lyValValAlaHisAlaCysAsnProSerAlaLeu-GlyValGlnGlyGlyAlaAspHis 221

Db 44 AsptYrArg 46

Search completed: February 8, 2006, 11:19:08
Job time : 39.7434 secs

GenCore version 5.1.7
Copyright (c) 1993 - 2006 Bioceleration Ltd.

OM nucleic - protein search, using frame_plus_n2p model

Run on: February 8, 2006, 10:23:31 ; Search time 41.5475 Seconds
(without alignments)
3060.024 Million cell updates/sec

Title: US-10-607-806-1-T9182_COPY_8700_9600

Perfect score: 1640

Sequence: 1 tgtctattataataaaat.....gagctcgaggcaggagtatc 901

Scoring table:

BLOSUM62
Xgapop 10.0, Xgapext 0.5
Ygapop 10.0, Ygapext 0.5
Fgapop 6.0, Fgapext 7.0
Delop 6.0, Delext 7.0

Searched: 2166443 seqs, 705528306 residues

Total number of hits satisfying chosen parameters: 4332886

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Command line parameters:

-MODEL=frame+n2p model -DEV=xlp
-Q=/abs/ABSSWEB_spool/WARTIN10607806/runat_08022006_092018_27658/app_query.fasta_1
-DB=Uniprot -QMT=fastan -SUFFIX=n2p.rup -MINMATCH=0.1 -LOOPCL=0 -LOOPEXT=0
-UNITS=bits -START=1 -END=1 -MATRIX=Blosum62 -TRANS=human40.cdi -LIST=45
-DOCALLIGN=200 -THR SCORE=pct -THR MAX=100 -THR MIN=0 -ALIGN=15 -MODE=LOCAL
-OUTFMT=ptc -NORM=ext -HEADSIZE=500 -MINLEN=0 -MAXLEN=200000000 -HOST=abs02p
-USER=WARTIN10607806 @CGN 1.1 694 @runat_08022006_092018_27658 -NCPU=6 -ICPU=3
-NMAP -NEG_SCORES=0 -WAIT -DSPELOCK=100 -LONGLOG -DEV_TIMEOUT=120
-WARN_TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5 -FGAPOP=6 -FGAPEXT=7
-YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database : Uniprot_05.80.*
1: uniprot_sprot.*
2: uniprot_trembl.*

Pred. NO. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
C 1	493.5	30.1	593	1 ALU7_HUMAN	P39194 homo sapien
C 2	471	28.7	593	1 ALU6_HUMAN	P39193 homo sapien
C 3	445	27.1	591	1 ALU8_HUMAN	P39195 homo sapien
C 4	437.5	26.7	587	1 ALU2_HUMAN	P39189 homo sapien
C 5	436	26.6	591	1 ALU1_HUMAN	P39188 homo sapien
C 6	427.5	26.1	587	1 ALU3_HUMAN	P39190 homo sapien
C 7	427.5	26.1	593	1 ALU7_HUMAN	P39194 homo sapien
C 8	422	25.7	585	1 ALU5_HUMAN	P39192 homo sapien
C 9	411	25.1	591	1 ALU1_HUMAN	P39198 homo sapien
C 10	400	24.4	585	1 ALU5_HUMAN	P39192 homo sapien
C 11	400	24.4	593	1 ALU6_HUMAN	P39193 homo sapien
C 12	394	24.0	591	1 ALU8_HUMAN	P39195 homo sapien
C 13	391.5	23.9	179	2 Q6ZS15_HUMAN	Q6ZS15 homo sapien
C 14	389.5	23.8	375	2 Q6O448_HUMAN	Q6O448 homo sapien
C 15	371.5	22.7	587	1 ALU2_HUMAN	P39189 homo sapien
C 16	368	22.4	164	2 Q6ZTS0_HUMAN	Q6ZTS0 homo sapien

C 17	366.5	22.3	156	2 Q8NBH4_HUMAN	Q8NBH4 homo sapien
C 18	361	22.0	146	2 Q6ZRX6_HUMAN	Q6ZRX6 homo sapien
C 19	360.5	22.0	128	2 Q6ZPB2_HUMAN	Q6ZPB2 homo sapien
C 20	359	21.9	239	2 Q6ZWA9_HUMAN	Q6ZWA9 homo sapien
C 21	350.5	21.4	129	2 Q6ZUW8_HUMAN	Q6ZUW8 homo sapien
C 22	348	21.2	170	2 Q8NAI3_HUMAN	Q8NAI3 homo sapien
C 23	344.5	21.0	150	2 Q6ZPA0_HUMAN	Q6ZPA0 homo sapien
C 24	344	21.0	123	2 Q6ZMQ3_HUMAN	Q6ZMQ3 homo sapien
C 25	342	20.9	129	2 Q8N9K0_HUMAN	Q8N9K0 homo sapien
C 26	339.5	20.7	587	1 ALU3_HUMAN	P39190 homo sapien
C 27	336	20.5	124	2 Q6ZSN0_HUMAN	Q6ZSN0 homo sapien
C 28	334	20.4	164	2 Q6ZUK0_HUMAN	Q6ZUK0 homo sapien
C 29	333	20.3	137	2 Q6ZS97_HUMAN	Q6ZS97 homo sapien
C 30	331.5	20.2	127	2 Q6ZVM2_HUMAN	Q6ZVM2 homo sapien
C 31	329	20.1	167	2 Q6ZSN3_HUMAN	Q6ZSN3 homo sapien
C 32	328.5	20.0	120	2 Q6ZTA3_HUMAN	Q6ZTA3 homo sapien
C 33	328	20.0	140	2 Q6ZP99_HUMAN	Q6ZP99 homo sapien
C 34	327.5	20.0	115	2 Q9N083_MACFA	Q9N083 macaca fasc
C 35	327	19.9	131	2 Q6ZTD3_HUMAN	Q6ZTD3 homo sapien
C 36	325	19.8	132	2 Q6ZS28_HUMAN	Q6ZS28 homo sapien
C 37	324	19.8	157	2 Q8N845_HUMAN	Q8N845 homo sapien
C 38	324	19.8	202	2 Q6ZUA4_HUMAN	Q6ZUA4 homo sapien
C 39	324	19.8	204	2 Q8N8P8_HUMAN	Q8N8P8 homo sapien
C 40	324	19.8	299	2 Q6ZRN6_HUMAN	Q6ZRN6 homo sapien
C 41	323.5	19.7	166	2 Q6ZUC0_HUMAN	Q6ZUC0 homo sapien
C 42	323	19.7	145	2 Q4R5J6_MACFA	Q4R5J6 macaca fasc
C 43	323	19.7	603	1 ALU4_HUMAN	P39191 homo sapien
C 44	322	19.6	132	2 Q6ZNU7_HUMAN	Q6ZNU7 homo sapien
C 45	321.5	19.6	139	2 Q6ZUL7_HUMAN	Q6ZUL7 homo sapien

ALIGNMENTS

RESULT 1

ALU7_HUMAN ID ALU7_HUMAN STANDARD; PRT; 593 AA.
AC P39194;
DT 01-FEB-1995 (Rel. 31, Created)
DT 01-FEB-1995 (Rel. 31, Last sequence update)
DT 10-MAY-2005 (Rel. 47, Last annotation update)
DE Alu subfamily S3 sequence contamination warning entry.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini; Hominidae;
OC Homo
OX NCBI_TaxID=9606;
RN [1]
RP NUCLEOTIDE SEQUENCE.
RX MEDLINE=95021758; PubMed=7935834; DOI=10.1038/371752a0;
RA Claverie J.-M., Makalowski W.;
RT "Alu alert";
RL Nature 371:752-752(1994).
RN [2]
RP CONCEPT.
RX MEDLINE=92241891; PubMed=1572661;
RA Claverie J.-M.;
RT "Identifying coding exons by similarity search: alu-derived and other potentially misleading protein sequences.";
RL Genomics 12:838-841(1992).
RN [3]
RP ALU FAMILIES CLASSIFICATION.
RX MEDLINE=88333009; PubMed=3138422;
RA Quentin Y.;
RT "The Alu family developed through successive waves of fixation closely connected with primate lineage history.";
RL J. Mol. Evol. 27:194-202(1988).
RN [4]
RP ALU FAMILIES CLASSIFICATION.
RX MEDLINE=91178815; PubMed=1706781;
RA Jurka J., Milosavljevic A.;
RT "Reconstruction and analysis of human Alu genes.";
RL J. Mol. Evol. 32:105-121(1991).
CC -!- MISCELLANEOUS: Various analyses indicate that Alu repeats fall

RP CONCEPT.
RX MEDLINE=92241891; PubMed=1572661;
RA Claverie J.-M.;
RT "Identifying coding exons by similarity search: alu-derived and other
RL potentially misleading protein sequences."; Genomics 12:838-841(1992).
RN (3)
RP ALU FAMILIES CLASSIFICATION.
RX MEDLINE=89333009; PubMed=3138422;
RA Quentin Y.;
RT "The Alu family developed through successive waves of fixation closely
RL connected with primate lineage history."; J. Mol. Evol. 27:194-202(1988).
RN (4)
RP ALU FAMILIES CLASSIFICATION.
RX MEDLINE=91178815; PubMed=1706781;
RA Jurka J., Milosavljevic A.;
RT "Reconstruction and analysis of human Alu genes."; J. Mol. Evol. 32:105-121(1991).
CC -I- MISCELLANEOUS: Various analyses indicate that Alu repeats fall
CC into 8 subfamilies (Ref.3 and Ref.4). Therefore, 8 Alu warning
CC consensus sequences have been constituted that contain all six
CC frames conceptual translations of each of these classes of Alu
CC repeats.
CC -I- MISCELLANEOUS: Isolated 'X' indicates the presence of a stop
CC codon. 'XXX' is used to separate the various translation phases.
CC -I- CAUTION: This Alu entry is provided in order to avoid the further
CC pollution of protein sequence databases with Alu-derived amino
CC acid sequences.
CC -I- CAUTION: Alu repetitive sequences are interspersed in human and
CC primate genomes with an average spacing of 4 kb. Some of them are
CC actively transcribed by pol III. Normal transcripts may contain
CC Alu-derived sequences in 5' or 3' untranslated regions. However,
CC cDNA libraries also contain partial and/or rearranged cDNAs
CC ligated with Alu-derived sequence in any orientation. Although Alu
CC elements (especially situated on the complementary strand) have a
CC great potential to create additional/alternative exons,
CC consideration should be given to the possibility that the presence
CC of an Alu in an open reading frame may have resulted from a
CC cloning artifact or may be due to misinterpretation of sequencing
CC data. This point has been overlooked on several occasions, with
CC the consequence of erroneous Alu-derived amino acid sequences
CC being reported.
CC -I- CAUTION: Any significant similarity of a putative protein sequence
CC with an Alu-translated entry must be taken as a warning that a
CC part of Alu repeat may have been artifactually included in the
CC coding nucleotide sequence.
CC -----
CC This Swiss-Prot entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use as long as its content is in no way modified and this statement is not
CC removed.
CC -----
CC EMBL; U14574; -; NOT ANNOTATED_CDS; Genomic_DNA.
CC REGION 1 96
CC REGION 100 195 Frame-1.
CC REGION 109 294 Frame-2.
CC REGION 199 294 Frame-3.
CC REGION 298 393 Frame-4.
CC REGION 397 492 Frame-5.
CC REGION 496 591 Frame-6.
CC SEQUENCE 591 AA; 64434 MW; AC8154D8A6BB280 CRC64;

CC Alignment Scores:
CC Pred. No.: 1.79e-40 Length: 591
CC Score: 445.00 Matches: 128
CC Percent Similarity: 46.1% Conservative: 25
CC Best Local Similarity: 38.6% Mismatches: 66
CC Query Match: 27.1% Indels: 113
CC DB: 1 Gaps: 10

CC US-10-607-806-1-T9182_COPY_8700_9600 (1-901) x ALU8_HUMAN (1-591)

Qy	901	GATACCTCCTGCTCAGACTCTGAGTAGCTGGGACTACAGGTGCTTGGCCACCATGCCAG	842
Db	329	AspSerProAlaSerAlaSerArgValAlaGlyIleThrGlyAlaArgHisHisAlaArg	348
Qy	841	CTAAATTTTGTATTTTATAGAGAGGGTGTTCATCATGTTGTCAGGCTGCTCAAA	782
Db	349	LeuIlePheValPheLeuValGluThrGlyPheHisValGlyGlnAlaGlyLeuGlu	368
Qy	781	CTCCGACCTCAAGTGCATTCATTTGCTGCGCTCCCAAGTCTCGGATACAGGCTGA	722
Db	369	LeuLeuThrSerGlyAsp-----LeuProProLysValLeuGlyLeuGlnAla***	385
Qy	721	GCCACCATGCCCAACCTA-----GTTTCTCTACTT	692
Db	386	AlaThrAlaProGlyLeuArgPro*****PhePhe***AspGlyValSerLeuCys	405
Qy	691	-----TAAAGCATGTATTTTAAATAATATGGAACATCTTAAATGGA-----	650
Db	406	ArgProGlyTrpSerAlaValAlaArgSerArgLeuThrAlaThrSerAlaSerArgVal	425
Qy	649	-----ATACTACATGACATTTAGAAAATTTAGTATGCTGCACATGGAATTTT	599
Db	426	GlnAlaIleLeuLeuProGlnProProGlu**LeuGlyLeuGlnAlaArgAlaThrThr	445
Qy	598	CCC-----	596
Db	446	ProGly***PheLeuTyrPhe*****ArgArgGlyPheThrMetLeuAlaArgLeuVal	465
Qy	595	-----AAATAAAAAAGCTAGTGTATGCCCAAAATCTGGAGGAATATGTAACGTGT	542
Db	466	SerAsnSer***ProGlnValIlePheLeuProLysCysTrp-----	479
Qy	541	ATCTTGGGTAGTAGGATATATAGATGCTTCCACTTCCATATAGATATATATGATAG	482
Db	480	-----AspTyrArgArgGluPro-----	485
Qy	481	TTTGAAGTTGCATTACTTTTGGCTTGTTCATTTTAAACCAAAATAACAATAAGATTA	422
Db	486	-----ProArg-----	487
Qy	421	CAGTAAATGSCATTTAAACTTTTATACATATAATTTACGTATCTCTGCTTACCTTCAGG	362
Db	487	-----	487
Qy	361	ATCTTTATAGTTTGTTCCTCCGATCTTTATTTTATTTATTTTATTTTGTAGATA	302
Db	488	-----ProAlaSerAlaArg***-----PhePheGluThr	499
Qy	301	GAGTCTCGCTGCTCGCCAGGCTGGAGTGCAGTGGTGGATCTCGGCTCACTGCAACCT	242
Db	500	GluSerArgSerValAlaGlnAlaGlyValGlnTrpArgAspLeuGlySerLeuGlnPro	519
Qy	241	CTGCCTCCCAAGGTTCAAGCAATTTCTTGCCTCAGCTCCCGAGTACGCTGGACTACAG	182
Db	520	ProProProGlyPheLysArgPheSerCysLeuSerLeuProSerSerTrpAspTyrArg	539
Qy	181	CATGCATCACTACACCAGCTAATTTTGTATTTTATAGTAGATAGGTTTACCCATAT	122
Db	540	ArgAlaProProArgProAlaAlaPheCysIlePheSerArgAspGlyValSerProCys	559
Qy	121	TGACCCAGGCTGCTCGAACTTCTCACCTCAAGTGTATCCACCACCTTGGACTTCCAAAG	62
Db	560	TrpProGlyTrpSerArgThrProAspLeuArg***SerSer-Ser-----GlnSe	576
Qy	61	TACTGTGGTTACAGGCTTGTAGTCACTGTGCCCAA	28
Db	576	rAlaGlyIleThrGlyValSerHisArgAlaArg	587
Db		RESULT 4	
ALU2_HUMAN		STANDARD;	
ID ALU2_HUMAN		PRT;	587 AA.
AC P39189;			


```
QY 71 ACTTCCAACTACTGTGTTACAGCGTTGAGTCACTGTGCC 30
Db 573 ProProLysValLeuGlyLeuGlnAla***AlaThrAlaPro 586

RESULT 5
ALU1_HUMAN STANDARD; PRT; 591 AA.
AC P39188; 1995 (Rel. 31, Created)
DT 01-FEB-1995 (Rel. 31, Last sequence update)
DT 10-MAY-2005 (Rel. 47, Last annotation update)
DE Alu subfamily J sequence contamination warning entry.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini; Hominidae;
OC Homo.
OX NCBI_TaxID=9606;
RN [1]
RP NUCLEOTIDE SEQUENCE.
RX MEDLINE=95021758; PubMed=7935834; DOI=10.1038/371752a0;
RA Claverie J.-M., Makalowski W.;
RT "Alu alert.";
RL Nature 371:752-752(1994).
RN [2]
RP CONCEPT.
RX MEDLINE=92241891; PubMed=1572661;
RA Claverie J.-M.;
RT "Identifying coding exons by similarity search: alu-derived and other
RT potentially misleading protein sequences.";
RL Genomics 12:838-841(1992).
RN [3]
RP ALU FAMILIES CLASSIFICATION.
RX MEDLINE=88333009; PubMed=3138422;
RA Quentin Y.;
RT "The Alu family developed through successive waves of fixation closely
RT connected with primate lineage history.";
RL J. Mol. Evol. 27:194-202(1988).
RN [4]
RP ALU FAMILIES CLASSIFICATION.
RX MEDLINE=91178815; PubMed=1706781;
RA Jurka J., Milosavljevic A.;
RT "Reconstruction and analysis of human Alu genes.";
RL J. Mol. Evol. 32:105-121(1991).
CC -!- MISCELLANEOUS: Various analyses indicate that Alu repeats fall
CC into 8 subfamilies (Ref.3 and Ref.4). Therefore, 8 Alu warning
CC consensus sequences have been constituted that contain all six
CC frames conceptual translations of each of these classes of Alu
CC repeats.
CC -!- MISCELLANEOUS: Isolated 'X' indicates the presence of a stop
CC codon, 'XXX' is used to separate the various translation phases.
CC -!- CAUTION: This Alu entry is provided in order to avoid the further
CC pollution of protein sequence databases with Alu-derived amino
CC acid sequences.
CC -!- CAUTION: Alu repetitive sequences are interspersed in human and
CC primate genomes with an average spacing of 4 kb. Some of them are
CC actively transcribed by pol III. Normal transcripts may contain
CC Alu-derived sequences in 5' or 3' untranslated regions. However,
CC cDNA libraries also contain partial and/or rearranged cDNAs
CC ligated with Alu-derived sequence in any orientation. Although Alu
CC elements (especially situated on the complementary strand) have a
CC great potential to create additional/alternative exons,
CC consideration should be given to the possibility that the presence
CC of an Alu in an open reading frame may have resulted from a
CC cloning artifact or may be due to misinterpretation of sequencing
CC data. This point has been overlooked on several occasions, with
CC the consequence of erroneous Alu-derived amino acid sequences
CC being reported.
CC -!- CAUTION: Any significant similarity of a putative protein sequence
CC with an Alu-translated entry must be taken as a warning that a
CC part of Alu repeat may have been artifactually included in the
CC coding nucleotide sequence.
CC -----
CC This Swiss-Prot entry is copyright. It is produced through a collaboration
```

```
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use as long as its content is in no way modified and this statement is not
CC removed.
CC -----
CC EMBL; U14567; -; NOT_ANNOTATED_CDS; Genomic_DNA.
CC -----
KW Hypothetical protein_1 96 Frame-1.
FT REGION 100 195 Frame-2.
FT REGION 199 294 Frame-3.
FT REGION 298 393 Frame-4.
FT REGION 397 492 Frame-5.
FT REGION 496 591 Frame-6.
SQ SEQUENCE 591 AA; 63827 MW; 665D395735519D95 CRC64;

Alignment Scores:
Pred. No.: 1,84e-39 Length: 591
Score: 436.00 Matches: 131
Percent Similarity: 47.9% Conservative: 18
Best Local Similarity: 42.1% Mismatches: 87
Query Match: 26.6% Indels: 75
DB: 1 Gaps: 7

US-10-607-806-1-T9182_COPY_8700_9600 (1-901) x ALU1_HUMAN (1-591)
QY 901 GATACCTCCTCCTCAGACTCCTGAGTAGCTGGGACTACAGTCTTGCACCATGCCAG 842
Db 329 AspProAlaSerAlaSerArgValAlaGlyThrGlyAlaArgHisHisAlaArg 348
QY 841 CTAATTTTGTATTTTATTAGAGAGGGTGTTCATCATCTGTCGCCAGGCTGCTCAAA 782
Db 349 LeuIlePheValPheValGluThrGlyPheHisTyrValAlaGlnAlaGlyLeuGlu 368
QY 781 CTCGGACCTCAAGTGATCATTTGCTCGGCTCCCAAGTCTGGGATACAGGCGTGA 722
Db 369 LeuLeuGlySerSerAspProAlaSerAlaSerGlnSerAlaGlyIle----- 385
QY 721 GCACACATGCCCAACCTAGTTTCTCTACTTTAAAGCATGTATTTTAAAAATATTGNAACA 662
Db 385 ----- 385
QY 661 TCCTTAAATGGAATACATACATAGACATACATAGAAAAATTAGTATGCTGCACATGGAAT 602
Db 386 -----ThrGlyValSerHisArg-----AlaArg*****Ph 397
QY 601 TTTCCTCAATAAAAAAGCTAGTGTATGCCCAAAATCTGGAGGATATGTACTGTT 542
Db 397 ePhe***AspArgValSerLeuCysArgPro-----GlyTrpSe 410
QY 541 ATCTTTGGGT---AGTAGGATTATAGATGCTTCCACTTTCATATATAGATATTATGTA 485
Db 410 rAlaValAlaArgSerArgLeuThrAlaAla-SerThr-SerArgAlaGln----- 426
QY 484 GAGTTTGAAGTTGCAATTACTTTTGGCTGTGTTTACTTTTATACCAAAATAACAATAAGA 425
Db 427 -----AlaIleLeuLeuProGlnProGlu***LeuGly 438
QY 424 TTACAGTTAAATGCAATTTAACTTTTATACATATATTTACGTATCTCTGCTTACCTTC 365
Db 439 LeuGlnAlaArgAlaThrThrProGly***PheLeuTyrPheLeu***ArgArgGlyPhe 458
QY 364 AGGAATCTTATTAGTTTGTCTCCGATCTCTTATTTTATTTTATTTTATTTTATTT 317
Db 459 ThrMetLeuProArgLeuValSer---AsnSerTrpAlaGlnValIleLeuProProArg 477
QY 316 -----TATTTTATTTT 308
Db 478 ProProLysValLeuGlyLeuGlnAla***AlaThrAlaProGly*****PhePhe 497
QY 307 GAGATAGACTCTGCTGTGCGCCAGGCTGGAGTGCAGTGGTGGGATCTCGGCTCACTG 248
Db 498 GluThrGlySerArgSerValAlaGlnAlaGlyValGlnTrpArgAspHisGlySerLeu 517
```


Db 88 -----AsnLys---SerGluThrProSerGlnLysLys*****----- 100
 QY 392 ATGTAATAAGTTTAAATGCAATTTTAACTGTAATCTTATTGTTTATTTGTTTATAAAG 451
 Db 100 ----- 100
 QY 452 TAAACAAGCCAAAGTAAGTCAACTTCAAACTCTACATAAATCTATTATGGAAGTGG 511
 Db 100 ----- 100
 QY 512 AAGCATCTATAATCCTACTACTACCAAGATAACAGGTACATATCTCTCCAGATTTTGG 571
 Db 100 ----- 100
 QY 572 GGCATACACTAGCTTTTATTGCGAATAATTTCCATGTGCGAGCATACCTAATTTTC 631
 Db 100 ----- 100
 QY 632 TAAATGCTATAGTAGTATCCATTAAGGATGTTCCATAATTTTAAAAATACATGCTTA 691
 Db 100 ----- 100
 QY 692 AAGTAGAAGAACTAGTTGGCATGCGTTCAGCTCTGTA-TCCAGAGCACTTTGGAGG 750
 Db 101 -----AlaGlyArgGlyGlySerArgLeu***SerGlnHisPheGlyArg 115
 QY 751 CCGAGGCAATGGATCACTTGGAGTCCGGAGTTTGAGACACGAGCTGGACACATGATGAA 810
 Db 116 ProArgValPheHisLeuArgSerGlyValArgAspGlnProGlyGlnHisGlyGlu 135
 QY 811 ACACCTCTCTTAAATAAATACAAAATTTAGCTGGCATGTTGGCAGCAGCTGTAGTCC 870
 Db 136 ThrProSerLeuLeuLysLeuAlaGlyArgGlyGlyArgLeu***Ser 155
 QY 871 CAGCTACTCAGGAGTCTGAGCAGGAG 897
 Db 156 GlnLeuLeuGlyArgLeuArgGlnGlu 164

RESULT 8
 ID ALU5_HUMAN STANDARD; PRT; 585 AA.
 AC P39192;
 DT 01-FEB-1995 (Rel. 31, Created)
 DT 01-FEB-1995 (Rel. 31, Last sequence update)
 DT 10-MAY-2005 (Rel. 47, Last annotation update)
 DE Alu subfamily SC sequence contamination warning entry.
 OS Homo sapiens (human).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini; Hominidae;
 OC Homo.
 OX NCBI_TaxID=9606;
 RN [1]
 RP NUCLEOTIDE SEQUENCE.
 RX MEDLINE=95021758; PubMed=7935834; DOI=10.1038/371752a0;
 RA Claverie J.-M., Makalowski W.;
 RT "Alu alert";
 RL Nature 371:752-752(1994).
 RN [2]
 RP CONCEPT.
 RX MEDLINE=92241891; PubMed=1572661;
 RA Claverie J.-M.;
 RT "Identifying coding exons by similarity search: alu-derived and other
 RT potentially misleading protein sequences";
 RL Genomics 12:838-841(1992).
 RN [3]
 RP ALU FAMILIES CLASSIFICATION.
 RX MEDLINE=88333009; PubMed=3138422;
 RA Quentin Y.;
 RT "The Alu family developed through successive waves of fixation closely
 RT connected with primate lineage history";
 RL J. Mol. Evol. 27:194-202(1988).
 RN [4]
 RP ALU FAMILIES CLASSIFICATION.

RX MEDLINE=91178815; PubMed=1706781;
 RA Jurka J., Milosavljevic A.;
 RT "Reconstruction and analysis of human Alu genes";
 RL J. Mol. Evol. 32:105-121(1991).
 CC -1- MISCELLANEOUS: Various analyses indicate that Alu repeats fall
 CC into 8 subfamilies (Ref.3 and Ref.4). Therefore, 8 Alu warning
 CC consensus sequences have been constituted that contain all six
 CC frames conceptual translations of each of these classes of Alu
 CC repeats.
 CC -1- MISCELLANEOUS: Isolated 'X' indicates the presence of a stop
 CC codon, 'XXX' is used to separate the various translation phases.
 CC -1- CAUTION: This Alu entry is provided in order to avoid the further
 CC pollution of protein sequence databases with Alu-derived amino
 CC acid sequences.
 CC -1- CAUTION: Alu repetitive sequences are interspersed in human and
 CC primate genomes with an average spacing of 4 kb. Some of them are
 CC actively transcribed by pol III. Normal transcripts may contain
 CC Alu-derived sequences in 5' or 3' untranslated regions. However,
 CC cDNA libraries also contain partial and/or rearranged cDNAs
 CC ligated with Alu-derived sequence in any orientation. Although Alu
 CC elements (especially situated on the complementary strand) have a
 CC great potential to create additional/alternative exons,
 CC consideration should be given to the possibility that the presence
 CC of an Alu in an open reading frame may have resulted from a
 CC cloning artifact or may be due to misinterpretation of sequencing
 CC data. This point has been overlooked on several occasions, with
 CC the consequence of erroneous Alu-derived amino acid sequences
 CC being reported.
 CC -1- CAUTION: Any significant similarity of a putative protein sequence
 CC with an Alu-translated entry must be taken as a warning that a
 CC part of Alu repeat may have been artifactually included in the
 CC coding nucleotide sequence.
 CC -----
 CC This Swiss-Prot entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use as long as its content is in no way modified and this statement is not
 CC removed.
 CC -----
 DR EMBL; U14571; -; NOT_ANNOTATED_CDS; Genomic_DNA.
 KW Hypothetical protein.
 FT REGION 1 95 Frame-1.
 FT REGION 99 193 Frame-2.
 FT REGION 197 291 Frame-3.
 FT REGION 295 389 Frame-4.
 FT REGION 393 487 Frame-5.
 FT REGION 491 585 Frame-6.
 SQ SEQUENCE 585 AA; 63994 MW; 46E8C4F493650A7 CRC64;
 Alignment Scores:
 Pred. No.: 6,92e-38 Length: 585
 Score: 422.00 Matches: 113
 Percent Similarity: 41.7% Conservative: 8
 Best Local Similarity: 39.0% Mismatches: 36
 Query Match: 25.7% Indels: 134
 DB: 1 Gaps: 1
 US-10-607-806-1-T9182_COPY_8700_9600 (1-901) x ALU5_HUMAN (1-585)
 QY 899 TACTCTCGCTCAGACTCTCTGAGTAGCTGGGACTACAGGTGCTGCACCATCCAGCT 840
 Db 424 PheSerCysLeuSerLeuProSerTrpAspTyrArgAlaProArgProAla 443
 QY 839 AATTTTGTATTTTATTATGAGAGGGTGTTCATCATGTTTCCAGGCTGGTCTCAACT 780
 Db 444 AsnPheCysIlePheSerArgAspGlyValSerProCysTrpProGlyTrpSerArgSer 463
 QY 779 CCGGACCTCAAGTATCCATTTCCTCGGCTCCCAAGTCTGGGATA-CAGGCTGAG 721
 Db 464 LeuAspLeu--ValIleArgProProArgProProLysValLeuGlyLeuGlnAla***A 483
 QY 720 CCACCATGCCCAACACTAGTTTCTCTACTTTAAAGCATGTATTTTAAATAATTTATGGAACAT 661

Db 23 GluProArgSerSerArgProAlaThr*****AsnProValSerThrLysAsn 42
QY 152 ACAAATAATAGCTGGTGTAGTATGATGCTAGTCCAGCTACTCGGAGCGCTGAG 211
Db 43 ThrIysIleSerArgAlaThrTTPArgAlaProValProAlaThrArgGluAlaGlu 62
QY 212 GCAAGAGAAATGCTTGAACCTGGAGGAGGAGGTGGAGTGGAGCGGAGATCCCACTG 271
Db 63 AlaGlySerLeuGluProGlyArgSerArgLeuGln**AlaValIleAlaProLeu 82
QY 272 CACTCCAGCTGGCGGACACAGCAGACTCTATCTCAAAAAAATAATAATAATAA 331
Db 83 HisSerSerLeuGlyArgAlaArgProCysLeuIlys***** 99
QY 332 GGATCGGAGAGAAACAAACTAATAAGATTCTCTGAAGGTAAAGCAGATACGTAAATAT 391
Db 99 ----- 99
QY 392 ATGTAATAAGTTTAAATGATTTTAACTGTAATCTTATTGTTTATTTGTTATATAAAG 451
Db 99 ----- 99
QY 452 TAAACAGCCAAAGTAATGCACTTCAACTCTACATAATATCTATTATGGAAGTGG 511
Db 99 ----- 99
QY 512 AAGGCATCTATAATCTTACTACTACCAAGATAACCAAGTTACATATCTCTCCAGATTTTGG 571
Db 99 ----- 99
QY 572 GGCATACACTAGCTTTTATTTTGGGAAATTTCCATGTGCAGGCATACCTAATTTTC 631
Db 99 ----- 99
QY 632 TAAATGCTATGATGATTCCATTAGGATGTTCCATATTTTAAATACATGCTTTA 691
Db 99 ----- 99
QY 692 AAGTAGAGAACTAGGTGGCTGGCTGCTCAGCTGTA-TCCAGCACTTTGGGAGG 750
Db 100 -----AlaGlyArgGlyGlySerArgLeu**SerGlnHisPheGlyArg 114
QY 751 CCGAGGCAATGATCATCTTGGAGTCCGGAGTTTGAGACCAAGCTGCGACACATGATGA 810
Db 115 ProArgArgGluAspHisLeuSerProGlyValArgAspGlnProGlyGlnHisSerGlu 134
QY 811 ACACCTCTCTAATAAATAACAATAATAGCTGGGATGCTGGCAGCAGCAGCTGTAGTCC 870
Db 135 ThrProSerLeuGlnIlysIleGlnLysLeuAlaGlyArgGlyAlaArgLeu***Ser 154
QY 871 CAGCTACTCAGGAGTCTGAGGCAGGAG 897
Db 155 GlnLeuLeuGlyArgLeuArgGlnGlu 163

RESULT 10

ALUS_HUMAN STANDARD; PRT; 585 AA.
ID ALUS_HUMAN
AC P39152;
DT 01-FEB-1995 (Rel. 31, Created)
DT 01-FEB-1995 (Rel. 31, Last sequence update)
DT 10-MAY-2005 (Rel. 47, Last annotation update)
DE Alu subfamily SC sequence contamination warning entry.
OS Homo sapiens (Human)
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini; Hominidae;
OC Homo.
OX NCBI_TaxID=9606;
RN [1]
RP NUCLEOTIDE SEQUENCE.
RX MEDLINE=95021758; PubMed=7935834; DOI=10.1038/371752a0;
RA Claverie J.-M., Makalowski W.;
RT "Alu alert."
RL Nature 371:752-752 (1994).

[2]
CONCEPT.
RP MEDLINE=92241891; PubMed=1572661;
RA Claverie J.-M.;
RT "Identifying coding exons by similarity search: alu-derived and other potentially misleading protein sequences.";
RL Genomics 12:838-841 (1992).
[3]
ALU FAMILIES CLASSIFICATION.
RP MEDLINE=88333009; PubMed=3138422;
RA Quentin Y.;
RT "The Alu family developed through successive waves of fixation closely connected with primate lineage history.";
RL J. Mol. Evol. 27:194-202 (1988).
[4]
ALU FAMILIES CLASSIFICATION.
RP MEDLINE=91178815; PubMed=1706781;
RA Jurka J., Milosavljevic A.;
RT "Reconstruction and analysis of human Alu genes.";
RL J. Mol. Evol. 32:105-121 (1991).
CC -I- MISCELLANEOUS: Various analyses indicate that Alu repeats fall into 8 subfamilies (Ref.3 and Ref.4). Therefore, 8 Alu warning consensus sequences have been constituted that contain all six frames conceptual translations of each of these classes of Alu repeats.
CC -I- MISCELLANEOUS: Isolated 'X' indicates the presence of a stop codon, 'XX' is used to separate the various translation phases.
CC -I- CAUTION: This Alu entry is provided in order to avoid the further pollution of protein sequence databases with Alu-derived amino acid sequences.
CC -I- CAUTION: Alu repetitive sequences are interspersed in human and primate genomes with an average spacing of 4 kb. Some of them are actively transcribed by pol III. Normal transcripts may contain Alu-derived sequences in 5' or 3' untranslated regions. However, cDNA libraries also contain partial and/or rearranged cDNAs ligated with Alu-derived sequence in any orientation. Although Alu elements (especially situated on the complementary strand) have a great potential to create additional/alternative exons, consideration should be given to the possibility that the presence of an Alu in an open reading frame may have resulted from a cloning artifact or may be due to misinterpretation of sequencing data. This point has been overlooked on several occasions, with the consequence of erroneous Alu-derived amino acid sequences being reported.
CC -I- CAUTION: Any significant similarity of a putative protein sequence with an Alu-translated entry must be taken as a warning that a part of Alu repeat may have been artifactually included in the coding nucleotide sequence.

CC This Swiss-Prot entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL Outstation - the European Bioinformatics Institute. There are no restrictions on its use as long as its content is in no way modified and this statement is not removed.

CC EMBL; U14571; -; NOT_ANNOTATED_CDS; Genomic_DNA.
KW Hypothetical protein.
FT REGION 1 95 Frame-1.
FT REGION 99 193 Frame-2.
FT REGION 197 291 Frame-3.
FT REGION 295 389 Frame-4.
FT REGION 393 487 Frame-5.
FT REGION 491 585 Frame-6.
SQ SEQUENCE 585 AA; 63994 MW; 46E8C4F493650A7 CRC64;

Alignment Scores:
Pred. No.: 2,07e-35 Length: 585
Score: 400.00 Matches: 109
Percent Similarity: 42.1% Conservative: 14
Best Local Similarity: 37.3% Mismatches: 36
Query Match: 24.4% Indels: 134
DB: 1 Gaps: 1

SQ SEQUENCE 593 AA; 64640 MW; 136EF344AACD12A2 CRC64;

Alignment Scores:

Pred. No.: 2,07e-35 Length: 593
Score: 400.00 Matches: 109
Percent Similarity: 42.2% Conservative: 13
Best Local Similarity: 37.7% Mismatches: 40
Query Match: 24.4% Indels: 128
DB: 1 Gaps: 3

US-10-607-806-1-t9182_COPY_8700_9600 (1-901) x ALU6_HUMAN (1-593)

Qy 32 GCACAGTACTCAAGCCTGTAAACACAGTACTTTGGAGTCCAAAGTGGGTGGATCACTT 91
Dy 3 AlaArgTrpLeuThrProAlaLeuTrpGluAlaGluAlaGlyGlySerPro 22
Qy 92 GAGGTGAGAGTTCGACGACGCTGTGTCAATATGTTGAAACCTCTCTACTAAAAAT 151
Dy 23 GluValGlySerSerArgProAla**ProThrTrpArgAsnProValSerThrLysAsn 42
Qy 152 AAAAAAATTAGCTGGGTGTAGTGTATGATGCTGTAGTCCAGCTACTCGGAGGCTGAG 211
Dy 43 ThrLysIleSerArgAlaTrpTrpArgMetProValIleProAlaThrArgGluAlaGlu 62
Qy 212 GCAAGAGAAATGCTTGAACCTGGAGGAGGAGTTCAGTGCAGCGAGATCCCACTG 271
Dy 63 AlaGlyGlySerLeuGluProGlyArgArgArgLeuArg**AlaGluIleAlaProLeu 82
Qy 272 CACTCCAGCTGGGCGACACGAGGACTCTATCTCAAAAAATAAATAATAATAATAA 331
Dy 83 HisSerSerLeuGly 87
Qy 332 GGATCGGAGAGAAACAAACCTAATAAGATTCCTGAAGTAAAGCAGAGATAGTAAAT 391
Dy 88 -----AsnLys-----SerGluThrProSerGlnLysLys***** 100
Qy 392 ATGTAATAAGTTTAAATGCAATTTAACTGTAATCTTTATTGTTTATTTGTTATAAAG 451
Dy 100 ----- 100
Qy 452 TAAACAGCAAAAGTAATGCAACTTCAAACTCTACATAATATCTATTATGAAAGTGG 511
Dy 100 ----- 100
Qy 512 AAGCATCTATATCTTACTACTACCAAGATAACCAAGTTACATATTCCTCCAGATTTTGG 571
Dy 100 ----- 100
Qy 572 GGCATACATAGCTTTTTTTTATTGGGAAATTTCCATGTGCAGGCATACCTAATTTTC 631
Dy 100 ----- 100
Qy 632 TAAATGCTATGTAGTATTCCATTTAGGATGTTCCATAATTTTAAATACATGCTTTA 691
Dy 100 ----- 100
Qy 692 AAGTAGAGAAATAGGTTGGCATGGTGGCTCAGCCTGTA-TCCAGCACTTTGGAGG 750
Dy 101 -----AlaGlyArgGlyGlySerArgLeu**SerGlnHisLysPheGlyArg 115
Qy 751 CCGAGGCAAAATGGATCACTTGGAGTCCGGAGTTTGAGACCAAGCCTGGACAAATGATGAA 810
Dy 116 ProArgArgAlaAspHisLeuArgSerGlyValArgAspGlnProAspGlnHisGlyGlu 135
Qy 811 ACACCTCTCTAATAAATAACAAATATAGCTGGGATGGTGGCAGACCTGTGATGTC 870
Dy 136 ThrProSerLeuLeuLysIleGlnLysLeuAlaGlyArgGlyGlyAlaCysLeu**Ser 155
Qy 871 CAGCTACTCAGGAGTCTGAGGAGGAG 897
Dy 156 GlnLeuLeuGlyArgLeuArgGlnGlu 164

RESULT 12

ALU6_HUMAN

ID ALU6_HUMAN STANDARD; PRT; 591 AA.
AC P39155.
DT 01-FEB-1995 (Rel. 31, Created)
DT 01-FEB-1995 (Rel. 31, Last sequence update)
DT 10-MAY-2005 (Rel. 47, Last annotation update)
DE Alu subfamily SX sequence contamination warning entry.
OS Homo sapiens (Human)
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini; Hominidae;
OC Homo.
OX NCBI_TaxID=9606;
RN [1]
RP NUCLEOTIDE SEQUENCE.
RX MEDLINE=95021758; PubMed=7935834; DOI=10.1038/371752a0;
RA Claverie J.-M., Makalowski W.;
RT "Alu alert."
RL Nature 371:752-752 (1994).
RN [2]
RP CONCEPT.
RX MEDLINE=92241891; PubMed=1572661;
RA Claverie J.-M.;
RT "Identifying coding exons by similarity search: alu-derived and other
potentially misleading protein sequences";
RL Genomics 12:838-841 (1992).
RN [3]
RP ALU FAMILIES CLASSIFICATION.
RX MEDLINE=88333009; PubMed=3138422;
RA Quentin Y.;
RT "The Alu family developed through successive waves of fixation closely
connected with primate lineage history";
RL J. Mol. Evol. 27:194-202 (1988).
RN [4]
RP ALU FAMILIES CLASSIFICATION.
RX MEDLINE=91178815; PubMed=1706781;
RA Jurka J., Milosavljevic A.;
RT "Reconstruction and analysis of human Alu genes";
RL J. Mol. Evol. 32:105-121 (1991).
CC -I- MISCELLANEOUS: Various analyses indicate that Alu repeats fall
into 8 subfamilies (Ref.3 and Ref.4). Therefore, 8 Alu warning
consensus sequences have been constituted that contain all six
frames conceptual translations of each of these classes of Alu
repeats.
CC -I- MISCELLANEOUS: Isolated 'X' indicates the presence of a stop
codon, 'xxx' is used to separate the various translation phases.
CC -I- CAUTION: This Alu entry is provided in order to avoid the further
pollution of protein sequence databases with Alu-derived amino
acid sequences.
CC -I- CAUTION: Alu repetitive sequences are interspersed in human and
primate genomes with an average spacing of 4 kb. Some of them are
actively transcribed by pol III. Normal transcripts may contain
Alu-derived sequences in 5' or 3' untranslated regions. However,
cDNA libraries also contain partial and/or rearranged cDNAs
ligated with Alu-derived sequence in any orientation. Although Alu
elements (especially situated on the complementary strand) have a
great potential to create additional/alternative exons,
consideration should be given to the possibility that the presence
of an Alu in an open reading frame may have resulted from a
cloning artifact or may be due to misinterpretation of sequencing
data. This point has been overlooked on several occasions, with
the consequence of erroneous Alu-derived amino acid sequences
being reported.
CC -I- CAUTION: Any significant similarity of a putative protein sequence
with an Alu-translated entry must be taken as a warning that a
part of Alu repeat may have been artifactually included in the
coding nucleotide sequence.

CC This Swiss-Prot entry is copyright. It is produced through a collaboration
between the Swiss Institute of Bioinformatics and the EMBL Outstation -
the European Bioinformatics Institute. There are no restrictions on its
use as long as its content is in no way modified and this statement is not
removed.

EMBL; U14574; -, NOT_ANNOTATED_CDS; Genomic_DNA.
 KW Hypothetical protein.
 FT REGION 1 96 Frame-1.
 FT REGION 100 195 Frame-2.
 FT REGION 199 294 Frame-3.
 FT REGION 298 393 Frame-4.
 FT REGION 397 492 Frame-5.
 FT REGION 496 591 Frame-6.
 SQ SEQUENCE 591 AA; 64434 MW; AC8154AD8A6B280 CRC64;
 Alignment Scores:
 Pred. No.: 9,81e-35 Length: 591
 Score: 394.00 Matches: 113
 Percent Similarity: 40.9% Conservative: 6
 Best Local Similarity: 38.8% Mismatches: 39
 Query Match: 24.0% Indels: 133
 DB: 1 Gaps: 4
 US-10-607-806-1-T9182_COPY_8700_9600 (1-901) x ALU8_HUMAN (1-591)
 QY 31 GGACAGCTGACTCAAGCCCTGTAAACACAGTACTTTTGGAAAGTCCAAAGTGGGTGGATCACT 90
 DB 6 GlyValAlaAlaHisAlaCysAsnProSerThrLeu-----Gly-GlyArgSerPr 22
 QY 91 TGAGGTGAGAAGTTGAGACACCGCTGGTCAATATGGTGAACCTTACTTCTACTATAAAA 150
 DB 22 oGluValArgSerArgProAlaTrpProThrTrp***AenProValSerThrLysAs 42
 QY 151 TACAAAAATTAGCTGGGTGTAGTGCATGCTGTAGTCCCGCTACTCCGGAGGCTGA 210
 DB 42 nThrLysIleSerArgAlaTrpTrpArgAlaProValIleProAlaThrArgGluAlaGl 62
 QY 211 GGCAAGAGATTCCTTGAACCTGGGAGGAGAGGTTGCAGTGAGCCGAGATCCCAACCACT 270
 DB 62 uAlaGlyGluSerLeuGluProGlyArgArgLeuGln***AlaGluIleAlaProLe 82
 QY 271 GCATCCAGCTGGGCGACACAGCGAGCTCTATCTCAAAAAATAAATAATAATAAATAA 330
 DB 82 uHisSerSerLeuGlyAspArgAlaArgLeuArgLeuLysLys*****AlaGlyG 102
 QY 331 AGGATCGGAGAGAAACAAACTAATGAATTCCTGAGGTAAGCAGAGATAGTAATTA 390
 DB 102 yGlyArgAlaArg----- 106
 QY 391 TATGTAATAAGTTTAAATGCACTTTTAACTGTAATCTTATTTGTTATTTTGGTTATAAA 450
 DB 107 -----TrpLeu----- 108
 QY 451 GTAAACAAGCCAAAAGTAATGCAACTTCAAACTCTACATAATATCTATTATGGAAGTG 510
 DB 108 ----- 108
 QY 511 GAAGGCATCTAATCTGTACTACCAAGATAACAGTTTACATATTCCTCAGATTGTTG 570
 DB 108 ----- 108
 QY 571 GGGCATACACTAGCTTTTATTTTGGGAAATTTCCATGTGCAGGCATACCTAATTTT 630
 DB 108 ----- 108
 QY 631 CTAATGTCCTATGTAGTATTTCCATTTTAAGGATGTTCCATAATTTTAAAAATACATGCTTT 690
 DB 108 ----- 108
 QY 691 AAAGTAGAGAAACTAGTTGGGCAATGGTGGCTCAGCCCTGTAATCCAGCATTTTGGGAGG 750
 DB 109 -----ThrProVal-IleProAlaLeuTrpGlu- 117
 QY 751 CCGAGGCAATGATCACTTGAGGTCGGAGTGTGAGACCGCTGGGACCATCATGAA 810
 DB 118 -----Glu-AspHisLeuArgSerGlyValArgAspGlnProGlyGlnHisGlyGlu 134
 QY 811 ACACCCCTCTCTAATAAAAAATACAAAAATTAAGCTGGGCATGGTGGCAAGCACCTGTAGTCC 870


```

Qy 752 CGAGCAAAATGGATCCTTGGAGTCCGGAGTTTGAGACCGCTGGACACATGATGAAA 811
Db 115 roArgAlaAspHis--GluValArgArgSerArgProSerTrpLeuThrArg***As 134
Qy 812 CACCTCTCTTAATAAAATACAAAATTAAGCTGGGCATGGTGGCAAGCACCTGTAGTCCC 871
Db 134 nProValSerThrLysAsnThrLysIleSerArgAlaTrpTrpArgAlaProValValPr 154
Qy 872 AGCTACTCAGGAGTCTGAGGCAGGA 896
Db 154 oAlaThrArgGluAlaGluAlaGly 162

```

Search completed: February 8, 2006, 11:14:23
Job time : 217.737 secs

THIS PAGE BLANK (USPTO)

GenCore version 5.1.7
Copyright (c) 1993 - 2006 Bioceleration Ltd.

OM nucleic - protein search, using frame_plus_n2p model

Run on: February 8, 2006, 10:22:47; Search time 70.7692 Seconds
(without alignments)
3105.549 Million cell updates/sec

Title: US-10-607-806-1-G7328-T9182_COPY_7000_9500
Perfect score: 4426
Sequence: 1 gtcgtgtcactgtgttcag.....tttgagaccagcctggacaa 2501

Scoring table: BLOSUM62

Xgapop 10.0, Xgapext 0.5
Ygapop 10.0, Ygapext 0.5
Fgapop 6.0, Fgapext 7.0
Delop 6.0, Delext 7.0

Searched: 2443163 seqs, 439378781 residues

Total number of hits satisfying chosen parameters: 4886326

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Command line parameters:
-MODEL=frame+ n2p.model -DEV=xlp
-O=/abs/ABSSWEB_spool/MARTIN10607806/runat_08022006_092016_27646/app_query.fasta_1
-DB=A_Geneseq -QFMT=fastan -SUFFIX=n2p.rag -MINMATCH=0.1 -LOOPCL=0 -LOOPEXT=0
-UNITS=bits -START=1 -END=1 -MATRIX=Blosum62 -TRANS=human40.cdi -LIST=45
-DOCALIGN=200 -THR SCORE=pct -THR MAX=100 -THR MIN=0 -ALIGN=15 -MODE=LOCAL
-OUTFMT=ptc -NORM=ext -HEAPSIZ=500 -MINLEN=0 -MAXLEN=200000000 -HOST=abs02p
-USER=MARTIN10607806 @CGN 1.1 605 @runat_08022006_092016_27646 -NCPU=6 -ICPU=3
-NOR MAP -NEG SCORES=0 -WAIT -DSPELOCK=100 -LONGLOG -DEV_TIMEOUT=120
-WARN_TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5 -FGAPOP=6 -FGAPEXT=7
-YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database : A_Geneseq 21:*

- 1: Geneseq1980s:*
- 2: Geneseq1990s:*
- 3: Geneseq2000s:*
- 4: Geneseq2001s:*
- 5: Geneseq2002s:*
- 6: Geneseq2003as:*
- 7: Geneseq2003bs:*
- 8: Geneseq2004s:*
- 9: Geneseq2005s:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
C 1	475	10.7	7067	6 AAE35265	Aae35265 Human P45
C 2	468.5	10.6	7052	6 AAE35266	Aae35266 Human P45
C 3	455.5	10.3	7052	6 AAE35266	Aae35266 Human P45
C 4	451	10.2	7036	6 AAE35264	Aae35264 Human P45
5	438.5	9.9	215	4 AAU31513	Aau31513 Novel hum
6	427.5	9.7	217	7 ADC87151	Adc87151 Human GPC
7	420.5	9.5	7067	6 AAE35265	Aae35265 Human P45
8	416	9.4	673	6 ADA84009	Ada84009 Human POM
9	413	9.3	185	4 AAO06014	Aao06014 Human pol

10	405	9.2	232	5 ADK36978	Adk36978 Novel hum
11	400	9.0	302	5 ADK36936	Adk36936 Novel hum
12	400	9.0	361	5 ABG68738	Abg68738 Human pro
13	400	9.0	449	5 ABJ01057	Abj01057 Ovary cel
14	397.5	9.0	304	4 AAU30391	Aau30391 Novel hum
15	397.5	9.0	375	3 AAB01399	Aab01399 Neuron-as
16	397.5	9.0	375	5 ABB81538	Abb81538 Neural th
17	397.5	9.0	375	5 AAE29142	Aae29142 AD7C-NTP
18	397.5	9.0	375	5 ABR63268	Abr63268 AD7C-NTP
19	397.5	9.0	375	6 ADA84017	Ada84017 Human POM
20	397.5	9.0	375	6 ABU03024	Abu03024 Human neu
21	397.5	9.0	375	6 ABB99774	Abb99774 Amino aci
22	397.5	9.0	375	6 ABJ19445	Abj19445 AD7C-neur
23	397.5	9.0	375	5 ADB37642	Adb37642 Human neu
24	397.5	9.0	375	8 ADR14409	Adr14409 Human NF-
25	397.5	9.0	375	9 AEA30020	Aea30020 Human neu
26	396.5	9.0	239	8 ADQ65417	Adq65417 Novel hum
27	391.5	8.8	673	6 ADA84009	Ada84009 Human POM
28	391.5	8.8	7036	6 AAE35264	Aae35264 Human P45
29	390	8.8	382	4 AAU32610	Aau32610 Novel hum
30	390	8.8	382	4 AAU31818	Aau31818 Novel hum
31	390	8.8	382	4 AAU32707	Aau32707 Novel hum
32	387.5	8.8	396	4 AAU30455	Aau30455 Novel hum
33	386	8.7	179	8 ADR09211	Adr09211 Human pro
34	384.5	8.7	158	7 ADB64893	Adb64893 Human pro
35	384	8.7	324	4 AAU29573	Aau29573 Novel hum
36	384	8.7	324	4 AAU29579	Aau29579 Novel hum
37	384	8.7	397	2 AAR59913	Aar59913 Neural th
38	383.5	8.7	257	4 AAO10294	Aao10294 Human pol
39	381	8.6	241	8 ADQ66812	Adq66812 Novel hum
40	381	8.6	527	7 ADD27133	Add27133 Human adi
41	380.5	8.6	175	4 AAU31857	Aau31857 Novel hum
42	380.5	8.6	250	4 AAU31823	Aau31823 Novel hum
43	380.5	8.6	250	7 ADE09441	Ade09441 Novel pro.
44	377	8.5	215	4 ABG07921	Abg07921 Novel hum
45	376	8.5	189	7 ADB65042	Adb65042 Human pro

ALIGNMENTS

RESULT 1

AAE35265

ID AAE35265 standard; protein; 7067 AA.

XX AAE35265;

XX AC

DT 28-MAY-2003 (first entry)

XX XX

DE Human P450RAI-3 protein #2.

XX XX

KW Human; retinoic acid; vitamin A; cytochrome P450; actinic keratosis;

KW P450RAI; cancer; psoriasis; acne; ichthyosis; gene therapy; vaccine;

KW enzyme.

XX XX

OS Homo sapiens.

XX XX

PN WO200295034-A2.

XX XX

PD 28-NOV-2002.

XX XX

PF 23-MAY-2002; 2002WO-CA000758.

XX XX

PR 23-MAY-2001; 2001US-0292531P.

XX XX

PA (CYTO-) CYTOCHROMA INC.

XX XX

PI Wisniewski J, Petkovich PM, Ramshaw H;

XX XX

DR WPI; 2003-112152/10.

XX XX

DR N-PSDB; AAD53861.

XX XX

PT New isolated P450RAI-3 nucleic acid molecule and polypeptide, useful for diagnosing, preventing, or treating disorders with aberrant expression or

Db 21 ValGluThrGlyPheHisCysIleSerGlnAspGlyLeuAspLeuMet---SerSer--S 39
Qy 849 CCTCTACTTGGCATCCCAAGTGTCTGGATTACAGGGGTGAGCCACCATGTCCGGCTA 908
Db 39 erThrLeuLeuGlyLeuProLysCysTrpAspTyrArgGluProArgCysProIlaG 59
Qy 909 CTTATTCTTTACATTCATCTTCCCAATAGAGTGAAGTCCAC----- 953
Db 59 ln-----GlnHisTyr--PheProLysIleAsnValHisGlnLeuMetIleGly 75
Qy 954 ---AGAACAGGATTTACTGCCTATTCTTT----- 980
Db 76 ThrArgCysGlyThrSerIleGlnTyrIleIleleCysGlnLysGlyMetLysCysMet 95
Qy 981 -----CCTTTCTTTTGGAGACAGAGTCTCATTTCATCACCCTCAACCTCC 1025
Db 96 LeuGlnArgAspGluProLysTyrTyrAlaValLysGlu--AlaSerTyrLysArgProG 115
Qy 1026 GTTCAGC----- 1032
Db 115 yArgSerPheCysPheValCysPheValLeuCysCysPheValLeuAsnArgGlyArgVa 135
Qy 1033 -----TCAGTCAACCTCTGCTCCCG 1055
Db 135 lLeuLeuCysCysProSerTrpSerThrValHisCysSerLeuGlnProProSerG 155
Qy 1056 GTTCAAGYGATTCCTCCCTTAAGCTCTGAGTAGCTGGAATTACAGAGGTGCACCA 1115
Db 155 yLeuLysLeuAlaSerHisIleuSerLeuProSer--HisTyrArgCysAlaProP 174
Qy 1116 ATGCTTCGGCTAATTTTGTATTTTACAGAGAGTGGGTTTACCAGTTGGCCAGGCT 1175
Db 174 oGlnProAlaAsnPheLeuTyr--Val--GluThrGlySerHisArgValAlaGlnAla 192
Qy 1176 GGTCTCAAACTCTGACCTCAAGTGTCTGCTGCTCCTCCAGTCTCCCAAGTGTGNA 1235
Db 193 GlyLeuLysLeuValSerSer----- 200
Qy 1236 ATAGGCTGAGTCACTGCTGCTGCGCGATTACTGTCTATTCTTTTATGCTATATCCC 1295
Db 201 LeuGlyLeuSerLysCysAspTyr----- 208
Qy 1296 AGATCTAGACGAGTGTGACATATAGTAGTGTCTCAATAAATGTAATGACACAG 1355
Db 209 -----ArgGlyGlu 211
Qy 1356 CCT-----AGATATAACTTCTTCTTTCTTTTCTTTTAAACAATCTTGACACT 1403
Db 212 ProProTrpAlaArgArgProHisValIlePheProPheGlyGluLeuLeuMetAlaThr 231
Qy 1404 -----TTGCAGATAAATACATCTT-----GCATTCTGCTTTTCACT 1442
Db 232 GlyPheLeuSerGlyAlaLeuLysLeuIleGluMetValAlaGlnPheCysGluSerGln 251
Qy 1443 TATCACCTTGTATGACTTTTTCATATGCTCAACCTTTATTTGTTACTGTTTTCAT 1502
Db 252 AsnHisPheValProPheLysValAsnCysMet--ValHisLeuLysAlaValLysLys 271
Qy 1503 TGTACTACTTTTACTGACTCAATAATATGCTTAATTTGTTTATATCATCTCTGCTCCA 1562
Db 271 le-----PheLysValIleGlnGluIleGluV 280
Qy 1563 CTTTAAAGCCCAAAATTTACAAATCTGATGAAGCTATGAACCTCTCCCCAGACAAATA 1622
Db 280 alLeuTyrSerTyrLysAsnLysAsnArgGlnLysTrpAspSerTyrProArgGluL 300
Qy 1623 CACACACACACACACTCTACACAGTTTTTTTTAAATGTTTTCGAATGAAGACAGAAA 1682
Db 300 euValSerLysValAsnSerHis--MetLeuAsnAsnLysLeuLysIleArgLysIle 319
Qy 1683 CTGCATTAGAGATGTTTGTTCATATTAATTAATAAATACTCAGTTGGGCACAGTGACT 1742
Db 319 erLeuLeuGlyIleValArgSerLeuGluLeuLysGluLeuLysAlaArgArgSerG 339

Qy 1743 CAAGCCTGTAAACACACAGTACTTTTGAAGTCCAAGGTGGTGGATCACTTTCAGGTGAGAAG 1802
Db 339 lySerHisLeuSerGlnHisPheGlyArgProArgTrpValAspHisLeuArgProGlyV 359
Qy 1803 TTCAGACACAGCCTGGTCAATATGGTGAACCCCTATCTCTACTATAAAATACAAAATTAG 1862
Db 359 al---AspGlnProGlyGlnHisGlyGluThrProSerLeuGlnLysThrGlnLysLeuA 378
Qy 1863 CTGGGTCTAGTGTATGATGCTCTAGTCCAGTACTCCGAGGCTCGAGGCTGAGCAAGAGATT 1922
Db 378 laGlyArgGlyGlyMetCysLeu---SerGlnLeuLeuGlyArgLeuArgHisLysAsnC 397
Qy 1923 GCTTTGAACCTG----- 1933
Db 397 yLeuAsnLeuArgProGlyAlaValAlaHisAlaCysAsnProSerThrLeuGlyGlyG 417
Qy 1933 ----- 1933
Db 417 lnAspGlyTrpIleThrArgSerGlyValAspGlnProAspGlnHisGlyGluThrProS 437
Qy 1933 ----- 1933
Db 437 erLeuThrLysAsnThrLysAsnProGlyMetValAlaCysAlaCysAsnAlaSerTyrS 457
Qy 1933 ----- 1933
Db 457 erGlyGlyGlyArgArgThrAlaThrTrpGluValAlaValSerArgAspHisA 477
Qy 1933 ----- 1933
Db 477 laThrAlaLeuGlnProGlyGlnGlnSerGluThrProSerGlnLysLysLysLysL 497
Qy 1934 -----GGAGCAGAGGTGTCAGTGCAGCCGAGATCCCACTGCA 1973
Db 497 yLysAsnCysLeuAsnProGlyGlyGluGlyCysSerGluProArgLeuHisCys 517
Qy 1974 CTCAGCCTGGGCGACACAGCAGACTCTATCTCAAAAAATAATAATAATAATAAAGG 2033
Db 517 hrProValTrpValThrGluArgAspSerValSerLysGln-----GlnGlnLys 535
Qy 2034 ATCGAGAGAAACAAACTAATAAGA-----TTCC 2063
Db 535 hrGlnLysThrAsnLysGlyValLysHisGlyTyrPheValArgGlnThrValGlyPhe 555
Qy 2064 TGAAGTAGCAGACATACGTAATATATATATAA----- 2099
Db 555 etileCysSer--IleLeuLeuThrMetTyrLeuIleLysLeuLeuLeuLeuAsnProHis 574
Qy 2100 AAGTTTAAATGCTATTTAACTGTAATCTTATTGTTTATTGTTTATAAAGTAAACAAG 2159
Db 575 LysPheThrCysMetTyrLeuTyrAlaIleLeuAlaMetValValThrLysLysLys 594
Qy 2160 CCAAAAGTAATGCACTTCAAACTCTACATAATATCTATTATGGAAGTGAAGCATC 2219
Db 595 IleLysIleLysGln-----IleTrpMet-----Cys 603
Qy 2220 TATATCTCTACTACCCAAAGATAACGAGTTACATATCTCCACGATTTTGGGGCATACA 2279
Db 604 ValAsnProIleMet-----IleLeuLeuTyrSerSerPhePhe 617
Qy 2280 CTAGCTTTTTTTTATTTGGGAAATTTCCATGTGCGGCATACCTATTTTCTTAAATGTC 2339
Db 618 -----PheLysPheGlnThrHisLeuGluValIleAspTyr 629
Qy 2340 TATGTAGTATTC-----CATTTAAGGATGTTCCATAAATTTTAAAAATCATGCTTT 2390
Db 630 TyrIleSerLeuAlaLysAsnThrLeuArgThrPheTyr-----IleCysMet 645
Qy 2391 AAGTAGAGAAACTAGTTGGGCATGGTGGCTCACGCTGTA----- 2432
Db 646 TyrSerSerProGluSerTyrLeuLysArgLeuIleProLeuSerSerPheCysSerMet 665

QY 2433 -----TCCAGACATTTGGAGCGC 2453
 Db 666 LysCysSerGluLeuThrProSerProHisArgGlyGlySerGluArgIleGlyGluGln 685
 QY 2454 AGGCAATGGATCAC-----TTGAGGTCCGGAGTTTGAGACCGACCTCGGA 2498
 Db 686 ValGlnMetGlyHisTrpLysArgAlaGlyLeuGlyGlyAlaSerAspProGluGly 705

RESULT 4
 AAE35264
 ID AAE35264 standard; protein; 7036 AA.
 AC AAE35264;
 XX
 DT 28-MAY-2003 (first entry)
 XX
 DE Human P450RAI-3 protein #1.
 XX
 KW Human; retinoic acid; vitamin A; cytochrome P450; actinic keratosis;
 KW P450RAI; cancer; psoriasis; acne; ichthyosis; gene therapy; vaccine;
 KW enzyme.
 XX
 OS Homo sapiens.
 XX
 PN WO200295034-A2.
 XX
 PD 28-NOV-2002.
 XX
 PF 23-MAY-2002; 2002WO-CA000758.
 XX
 PR 23-MAY-2001; 2001US-0292531P.
 XX
 PA (CYTO-) CYTOCHROMA INC.
 XX
 PI Wisniewski J, Petkovich PM, Ramshaw H;
 XX
 DR WPI: 2003-112152/10.
 DR N-PSDB; RAD53861.
 XX
 PT New isolated P450RAI-3 nucleic acid molecule and polypeptide, useful for
 PT diagnosing, preventing, or treating disorders with aberrant expression or
 PT activity of the P450RAI-3, such as cancer, actinic keratosis, acne and
 PT psoriasis.
 XX
 PS Example 1; Fig 1; 231pp; English.
 CC
 CC The invention relates to retinoic acid metabolising cytochrome P450,
 CC P450RAI polypeptides and polynucleotides. Methods and compositions of the
 CC invention are useful for diagnosing, preventing, ameliorating and/or
 CC treating disorders associated with the aberrant expression or activity of
 CC the P450RAI such as diseases related to vitamin A and retinoic acid
 CC metabolism, e.g. cancer, actinic keratosis, psoriasis, acne and
 CC ichthyosis. The invention is useful in gene therapy and as vaccine. The
 CC present sequence is human P450RAI protein
 SQ Sequence 7036 AA;
 Alignment Scores:
 Pred. No.: 1.15e-42 Length: 7036
 Score: 451.00 Matches: 238
 Percent Similarity: 38.6% Conservative: 107
 Best Local Similarity: 26.7% Mismatches: 238
 Query Match: 10.2% Indels: 312
 DB: 6 Gaps: 39

US-10-607-806-1-G7328-T9182_COPY_7000_9500 (1-2501) x AAE35264 (1-7036)

QY 2466 GATCCATTGCTCGGCTCCCAAGTCTCGGAT-ACAGCGGTGAGCCACCATGCCCAA 2408
 Db 37 AspProSerLeuAlaSerGlnSerAlaGlyIleThrGlyValSerHisGlyAlaArg 56
 QY 2407 CCTAGTTTCTCTACTTTAAAGCATGTATTTTAAAAAATTATGGAACATCTTTAAATGGAAT 2348
 |||||

Db 57 ProSer-----SerIlelle----- 61
 QY 2347 ACTACATAGACATTTAGAAAAATTAGGTATGCTGCACATGGAATTTCCCAATAAAA 2288
 |||||
 Db 62 -----PheAsnSerGlnLys 66
 QY 2287 AAAGCTAGTGTATGCCCAAAATCTGGAGGAATATGTAACCTGGTTATCTCTGGGTAGTA 2228
 |||||
 Db 67 ArgMetSerIle-----AsnLeuAspLysGln----- 75
 QY 2227 GGATTATAGATGCTTCCACTTTCCATATAGATATTTATGTAGAGTTTGAAGTTGCATT 2168
 |||||
 Db 76 GlyValValHisPro----- 80
 QY 2167 ACTTTTGGCTTGTACTTTTATACCAAAAATAAACAATAAGATTACAGTTAAATGTCAT 2108
 |||||
 Db 81 ---TyrAsnIleLeuLeuPheValAsnLysLysGluSerAsnAsnAlaCysTyrAsnVal 99
 QY 2107 TTAACCTTTTATACATATAATTTTACGTATCTCTGCTTACCTTCAGGAATCTTATTAGTTT 2048
 |||||
 Db 100 MetAsn-LeuGluAsnIleMetLeuLysLysLysProValThrLysAspGlnValGlyLe 119
 QY 2047 TGTTCCTCTCCGATCCTTTTATTATT-----ATTATTATTTTGGAGATAGAGTCTCG 1994
 |||||
 Db 119 uPheValLeu-----PheValLeuPheCysValValLeuPheIleGluAlaGlySerCy 137
 QY 1993 CTGTGTGCGCCAGGCTGGAGTGCAGTGGTGGGATCTCGGCTACCTGCACCTCTGCTCC 1934
 |||||
 Db 137 sSerAlaAlaHisProGlyValGlnPhe-----Ile-AlaHisCysSerLeuHisLeu 155
 QY 1933 CAGGTTCAAGCAATTTCTTCTGCTCAGCTCCGAGTAGCTGGAGTACAGGATGCATC 1874
 |||||
 Db 155 euAspSerSerProPro---ThrSerAlaSerGlnValAlaSerThrThrGlyValHis 174
 QY 1873 ACTACACCCAGCTAATTTTGTATTTTATAGATAGATAGAGTGTTCACCATATTGACCAGG 1814
 |||||
 Db 174 iHisSerGlnLeuIlePheAsnCysMetSerLysGlnGlyLeu-----T 189
 QY 1813 CTGGTCTCGAATCTTCCACTCAAGTGTACCCACCTGGACTTCCAAAGTACTGTGG 1754
 |||||
 Db 189 hrValLeuProArgLeuValSerAsnSerSerGlnAlaLeuAlaSerGlnSerAlaArg 209
 QY 1753 TTACAGCTTGTAGTCACTGTGCCCAACTGAGTTATTTTAAATATATATGAACAAATCC 1694
 |||||
 Db 209 leThrGlyValSerHisHisGlyArgProGluAspHisMetLeuPheAspSerHisLeu- 228
 QY 1693 TCTAATGTCAGTTTCTTGTCTTAGTTGCAAAACATTAATAAATAAATCTGTGTGAGTGT 1634
 |||||
 Db 229 -----GlySerAspC 232
 QY 1633 GTCTGTGTGTATTTCTCTGGGGAGAGGTTTCATAGCTTTTCATGATTTGTAAATTTG 1574
 |||||
 Db 232 ys-----TrpLeuArgGlyPhePheGln-----G 240
 QY 1573 GCCTT---CTAAAGTGGAGCAGGAGGTATATAAGCAAAATAAG-----C 1532
 |||||
 Db 240 lLeuAsnLeuArgTrpLeuHisAsnSerValAsnArgValLysThrIleAspLeuCysP 260
 QY 1531 CATATTATTCAGTCACTAAATAATAGTAACATGAATAAAGTAAAGTTTTCAGG 1472
 |||||
 Db 260 roLeuAsnGluLeuThrValTrpTyrIleSerLysLeu-PheLysLysLysPheLeuLys 279
 QY 1471 CAATATCAAAAAGTCAATAACAAGGTGATAGTGAAGAGCAGATGCAAGATTGTATTATA 1412
 |||||
 Db 280 LeuTyr-----AsnLysLysArgPheAsn 287
 QY 1411 TTCTGCAAAAGTGTCAAGATTGTTTAAAAA-----GAAAAAGAA 1370
 |||||
 Db 288 TyrIleLysValIleLysThrLysIleAspLysAsnGlyArgIleAlaThrGlnGluGlu 307
 QY 1369 AGTTTATATCTA-----GGCTGTGCATTCATCAATTATTATTG 1331
 |||||
 Db 308 SerTrpTyrLysValArgIleLeuIleCysAspSerCysLysIleIleAsn----- 324

CC to express them are also useful for producing the proteins. The proteins
CC are useful in genetic vaccination, testing and therapy, and can be used
CC as nutritional supplements. They may be used to increase stem cell
CC proliferation; to regulate haematopoiesis; and in bone, cartilage, tendon
CC and/or nerve tissue growth or regeneration; immune suppression and/or
CC stimulation; as anti-inflammatory agents; and in treatment of leukaemias.
CC AAU29510-AAU33304 represent the amino acid sequences of novel human
CC secreted proteins of the invention
XX
SQ Sequence 215 AA;

Alignment Scores:
Pred. No.: 7.16e-42 Length: 215
Score: 438.50 Matches: 116
Percent Similarity: 59.9% Conservative: 14
Best Local Similarity: 53.5% Mismatches: 60
Query Match: 9.9% Indels: 28
DB: 4 Gaps: 3

US-10-607-806-1-G7328-T9182_COPY_7000_9500 (1-2501) x AAU31513 (1-215)

Qy 639 TCTGTCCAGGAGGATGAGCAGTGTGTGACCATAGGTGACGCTCAACCTCCT 698
Db 4 SerValThrGlnAlaGlyValGlnTrpCysAsnLeuGlySerLeuGlnProLeuPro 23
Qy 699 GAGCTCAAGGATCTGCTGAGCTCAGCTCCAGTACCTGGGACTAGCGGCGACCA 758
Db 24 ArgPheArg***PheSerCysLeuSerLeuSerSerTrpAspTrpArgArgPro 43
Qy 759 CCAGCTGGCTAATTAATAAATTTTGTGAGACTGGGTCTTACTAGTTGGCCAG 818
Db 44 ProArgProAlaAn-PheCysIlePheSerArgAspMetAlaPheThrLeuAlaAr 63
Qy 819 GCTTGTCTTAACCTCTGGCTTCAAGC-AATCTCTACCTTGGCAGTCCCAAGTGTGG 877
Db 63 gLeuValSerAsnSer***ProGlnCysAspProProThrSerAlaSerGlnSerAlaG 83
Qy 878 GATTACAGGGGTGAGCCACCATGTGGCGTACTTATTCTTTACATTCCTTCCCAT 937
Db 83 uilethrGlyValSerHisArgAlaTrpProValIleuSerSerProGlnProPhe----- 101
Qy 938 AGAATGTAAGATCCACAGAACAGGATTAATCTTCTCTCTCTCTCTCTCTCTCTCT 997
Db 102 -----PhePhePheAspMe 106
Qy 998 AGAGTCTCATTATCCTCACTCACTCACTCACTCACTCACTCACTCACTCACTCACT 1042
Db 106 tGluSerHisAlaIleThrGlnAlaGlyVal-GlnTrpArgHisLeuGlySerLeuGlnP 126
Qy 1043 CTCTGCCTCCCGGTTCAAGYGATTCTCTGCTTAAGCTCTCAAGCTCTGGAATTACA 1102
Db 126 roProProProMetPhelys***SerSerCysLeuSerLeuSerSerTrpAspTrpA 146
Qy 1103 AGCGTGCACCACTGCTGTGCTAATTTTGTATTTT---AGCAGAGATGGGGTTTAA 1159
Db 146 rGArgProProProArgProAlaAnPhePhe-ValPheLeu***ArgAspGlyValSer 165
Qy 1160 CCATGTTGCCAGGCTGTCTCAAACTCTCAAGTCAAGTCAAGTCAAGTCAAGTCAAGT 1219
Db 166 ProCys***ProGlyTrpSerArgSerProAspLeuValIleHis-ProProTrpSerPr 185
Qy 1220 CCAAGTGTCTGGAATTATAGCGGTGAGTCAGTCACTGTGCTGCGCG 1262
Db 185 olysSerAlaGlyIleThrGlyLeuSerHisCysAlaGlnPro 199

RESULT 6

ID ADC87151 standard; protein; 217 AA.

XX ADC87151;

AC ADC87151;

XX 01-JAN-2004 (first entry)

XX

DE Human GPCR protein SEQ ID NO:1604.
XX
KW human; GPCR; guanosine triphosphate-binding protein coupled receptor;
KW gene therapy.
XX
OS Homo sapiens.
XX
PN EP1270724-A2.
XX
PD 02-JAN-2003.
XX
XX 18-JUN-2002; 2002EP-00013517.
XX
PR 18-JUN-2001; 2001JP-00246789.
XX
PA (NAAD-) NAT INST ADVANCED IND SCI & TECHNOLOGY.
PA (ADSC-) CENT ADVANCED SCI & TECHNOLOGY INCUBATIO.
XX
XX Suwa M, Asai K, Akiyama Y, Aburatani H;
XX
XX WPI; 2003-315783/31.
XX
DR N-PSDB; ADC87150.
XX
XX
PT New polynucleotide, useful for preparing a composition for treating a
PT patient in need of increased or suppressed activity or expression of the
PT guanosine triphosphate-binding protein coupled receptor.
XX
XX Claim 2; SEQ ID NO 1604; 28pp; English.
XX

CC The invention relates to a novel polynucleotide encoding a guanosine
CC triphosphate-binding protein coupled receptor (GPCR). A polynucleotide of
CC the invention may have a use in gene therapy. The polynucleotide and
CC polypeptide are useful for preparing a composition for treating a patient
CC in need of increased or suppressed activity or expression of the
CC guanosine triphosphate-binding protein coupled receptor. The protein
CC sequences shown in ADC85549-ADC87617 represent GPCR's of the invention.
XX

SQ Sequence 217 AA;

Alignment Scores:
Pred. No.: 1.44e-40 Length: 217
Score: 427.50 Matches: 115
Percent Similarity: 52.3% Conservative: 11
Best Local Similarity: 47.7% Mismatches: 64
Query Match: 9.7% Indels: 51
DB: 7 Gaps: 5

US-10-607-806-1-G7328-T9182_COPY_7000_9500 (1-2501) x ADC87151 (1-217)

Qy 499 TGGGAT-----AGGTACAATTATCTCTCACTTAACAGATAAGAAACT 540
Db 17 TrpAspGlnSerSerValGlnPheArgValAsnPheLeuLeuValAsp----- 33
Qy 541 GAGGCTCAGAAAGCTGAGCTATTGTGCCAAGATCACACAGCTTGTAGTGGTGCACAGTTT 600
Db 34 -----PheThrLysMetLysAsnPheCysSerSerGluAsnIle 46
Qy 601 GGGTTTTTTTTTGTGTTTGTAGACACAGGGCTTGTCTCTGTCAACCCAGCATGAGCAC 660
Db 47 LeuPhePhePheLeu-----ArgArgSerLeuThrLeuLeuProArgLeuGluCys 63
Qy 661 AGTGGTGCAACCATAGGTCACTGAGCTCAACTCTCTGAGCTCAAGGGATCTGTGACC 720
Db 64 SerGlyThrIleSerAlaHisCysAsnLeuHisLeuProGlySerGlyAspSerHisThr 83
Qy 721 TCAGCCTCCCAAGTAGCTGGGACTACGAGCGTGCACACACACCGCTGGCTAATTAATAAAA 780
Db 84 SerAlaSerGlnValAlaGlyIleThrGlyAlaHisArgHisAlaTrpLeuIlePheVal 103
Qy 781 ATTTTGTGTAGACTGGGTCTTACTACCTGGCCAGGCTTGTCTTAAACTCTGCGCTT 840
Db 104 --PheLeuLeuGluThrGlyPheCysHisValGlyGlnAlaGlyLeuGluLeuThrS 123

Db 441 nAla***AlaThrThrProSer*****PheValPheGlyGlyPhePhePh 461
Qy 921 CATTCCTCA----- 927
Db 461 ePheAlaLeuPheLeuPheLeuArg***AlaLeuAlaLeuThrProArgLeuGluCysSe 481
Qy 928 ----TCTTTCCAAATAGAAATGTAAGATCCACA-----GAACAGAGGATTACTGCCTAT 974
Db 481 rGlyLysPheTrpLeuThrAlaAlaSerThrSerTrpValGlnAlaLeuLeuProLe 501
Qy 975 T-----TTCTCTCTTC 986
Db 501 uSerProVal***LeuGlyLeuGlnAlaTrpAlaAlaLeuProGly***PheMetTrpPh 521
Qy 987 TTTTTCAGACAGATCTCACT-----TCATCACTCAAA-- 1020
Db 521 e*****ArgHisSerPheThrMetLeuLeuArgLeuValSerAsnSer***ProGlnVa 541
Qy 1021 -----CCTCCG----- 1026
Db 541 lIleCysProGlyLeuProLysCysTrpAspArgArgGluProProHisProAl 561
Qy 1026 ----- 1026
Db 561 a*****LeuPheLeuGlyValPhePhePhePheLeuLeuCysPheCysPhe** 581
Qy 1027 -----TTGAGTCACTGCA 1040
Db 581 *AspArgProLeuLeu***HisProGlyTrpSerAlaValAlaSerPheGlySerLeuGl 601
Qy 1041 ACCTCTCCCTCCCGGGTCAAGGATCTCTGCTAAGCTCTGCTAGTGGGAATTA 1100
Db 601 nProGlnProGlyPheLysArgPheSerCysLeuLeuCysLeuProCysSerTrpAspTy 621
Qy 1101 CAAGCGTGCCACCACTGCTGGCTAATTTTGTATTTTGTAGCAGAGATGGGGTTTAC 1160
Db 621 rArgHisGlyProProPheLeuAlaAsn-LeuCysIlePheAsnArgAspThrValSerP 641
Qy 1161 CATGTTGCCCAGGCTGCTCAACTCTGACCTCAAGTATCTGCTGCTGCTGCTGCTGCTCC 1220
Db 641 rCys***SerGlyTrpSerGlnThrProAspLeuLys***SerAlaLeuLeuValSer 660
Qy 1221 CAAGTCTGGAATTTATAGCGTGAGTCAC 1250
Db 661 GlnSerAlaGlyMetThrGlyValSerHis 670

RESULT 9

AAO06014
ID AAO06014 standard; protein; 185 AA.
XX AAO06014;
AC AAO06014;
DT 06-NOV-2001 (first entry)
DE Human polypeptide SEQ ID NO 19906.
XX Human; cytokine; cell proliferation; cell differentiation; gene therapy;
KW vaccine; peptide therapy; stem cell growth factor; haematopoiesis;
KW tissue growth factor; immunomodulatory; cancer; leukaemia;
KW nervous system disorders; arthritis; inflammation.
XX Homo sapiens.
OS
XX W0200164835-A2.
FN
XX
PD 07-SEP-2001.
XX
PF 26-FEB-2001; 2001WO-US004927.
XX
PR 28-FEB-2000; 2000US-00515126.
PR 18-MAY-2000; 2000US-00577409.
XX

(HYSE-) HYSEQ INC.

PA XX
PI XX
XX Tang YT, Liu C, Drmanac RT;
DR WPI; 2001-514838/56.
DR N-PSDB; AAI85945.
XX
PT Isolated nucleic acids and polypeptides, useful for preventing diagnosing
PT and treating e.g. leukemia, inflammation and immune disorders.
XX
PS Claim 20; SEQ ID NO 19906; 1399pp + Sequence Listing; English.
XX
CC The invention relates to human polynucleotides (AAI79941-AAI93841) and
CC the encoded proteins (AAO00010-AAO13910) that exhibit activity relating to
CC cytokine, cell proliferation or cell differentiation or which may induce
CC production of other cytokines in other cell populations. The
CC polynucleotides and polypeptides are useful in gene therapy, vaccines or
CC peptide therapy. The polypeptides have various cytokine-like activities,
CC e.g. stem cell growth factor activity, haematopoiesis regulating
CC activity, tissue growth factor activity, immunomodulatory activity and
CC activin/inhibin activity and may be useful in the diagnosis and/or
CC treatment of cancer, leukaemia, nervous system disorders, arthritis and
CC inflammation. Note: The sequence data for this patent did not form part
CC of the printed specification, but was obtained in electronic format
CC directly from WIPO at ftp.wipo.int/pub/published_pct_sequences
XX
SQ Sequence 185 AA;
Alignment Scores:
Pred. No.: 6,98e-39 Length: 185
Score: 413.00 Matches: 98
Percent Similarity: 54.7% Conservative: 12
Best Local Similarity: 48.8% Mismatches: 49
Query Match: 9.3% Indels: 42
DB: Gaps: 3
US-10-607-806-1-G7328-T9182_COPY_7000_9500 (1-2501) x AAO06014 (1-185)
Qy 642 GTCACCCAGCATGAGCAGACAGTGGTCAACCATAGTACTGACGCTCAACCTCTCTGAG 701
Db 11 IleAlaGlnAlaGlyValGlnTrpCysAsnLeuSerLeuGlnProLeuProProArg 30
Qy 702 CTCACGGATCTGCTGACCTCAGCTCCCAAGTAGCTGGAGCTACCGAGCTGCACCA 761
Db 31 PheLysArgPheSerCysLeuSerLeuProSerSerTrpAspTrpArgHisValProPro 50
Qy 762 CQCCTGGCTAATAAAAAATTTTGTAGACATGGGTCTTACTACGTGGCCAGGCT 821
Db 51 ArgLeuAlaAsnPhePheValPheLeuValGluMetGlyPheLeuHisValGlyGlnAla 70
Qy 822 TGTCTTAACTCTGCTTCAAGCAATCTCTCTACC-----TTG 860
Db 71 SerLeuGluLeuLeuAlaSerGlyAspProProSerGlyAspProAspLeuArg***Leu 90
Qy 861 GCATCCCAAAGTGTGGGATTACAGGGGTGAGCCACCATGTGCGGTACTTATTTCTTTA 920
Db 91 AlaPheGlnSerLeuGlyIleThrGlyValThrHisHis-AlaGlyGlnIlePheLeuPh 110
Qy 921 CATTCCATCTTTTCCAATAGAATGTAAAGATCCACAGAACAGGGATTACTGCCTATTTC 980
Db 110 eCysLeuGlu----- 113
Qy 981 CTTTCTTTTGTGACAGACAGTCTCATCTCATCTCACTCACTCACTCCG----- 1026
Db 114 -----ThrGluSerProSerValAlaGlnValGlyValG 125
Qy 1027 -----TTGAGTCACTGCACTCTGCTCCCGGTTCAAGYGATTTCTCTGCTCT 1076
Db 125 nTrpCysAsnLeuGlySerLeuGlnProLeuProProGlySerGlu***PheSerCysLe 145
Qy 1077 AAGCCTCTGAGTAGCTGGAATTAACAAGCGTGCAACCATGCTGGCTTAATTTTGTGA 1136
Db 145 uGlyLeuProSerSerTrpAspTrpArgHisAlaLeuThrHisProAlaAsn-PheSer 165

QY 1137 TTTTACAGAGTGGGTTTACCATGTTGCCAGGCTGGTCTCAACTCCTGACCTC 1195
 DB 165 lePheSerArgApeGlyValSerProSerTrpProGlyTrpSerGlnMetProAspLeu 184

RESULT 10
 ADK36978
 ID ADK36978 standard; protein; 232 AA.
 AC ADK36978;
 DT 06-MAY-2004 (first entry)
 XX Novel human polypeptide SeqID9060.

XX antiarthritic; antiparkinsonian; neuroprotective; nootropic;
 KW immunosuppressive; cytostatic; antipsoriatic; antiinflammatory;
 KW antibacterial; antiviral; antifungal; antiparasitic; gene therapy;
 KW arthritis; Parkinson's; Alzheimer's; autoimmune disease; cancer;
 KW psoriasis; inflammatory bowel disease; infection; bacteria; virus;
 KW fungus; parasite; human.

XX Homo sapiens.

XX Key Location/Qualifiers
 XX Misc-difference 1..232
 FT /label= OTHER
 FT /note= "OTHER= All Xaa's in this sequence are unknown
 FT amino acids or the site of a stop codon within the DNA
 FT sequence"

XX WO200216439-A2.

XX 28-FEB-2002.

XX 05-MAR-2001; 2001WO-US004941.

XX 07-MAR-2000; 2000US-00519705.

XX 19-MAY-2000; 2000US-00574454.

XX (HYSE-) HYSEQ INC.

XX Tang YT, Liu C, Drmanac RT;

XX WPI; 2002-280918/32.

XX Isolated polynucleotide encoding bone marrow derived polypeptides useful
 FT for treating, e.g., Parkinson's, Alzheimer's, cancer, arthritis, Crohn's
 FT disease, and inflammatory bowel disease.

XX Claim 20; SEQ ID NO 9060; 504pp; English.

XX This invention relates to a novel isolated polynucleotide comprising a
 CC nucleotide sequence selected from one of 1680 sequences, a mature protein
 CC coding portion of them, an active domain of them and their complementary
 CC sequences. The invention may be useful for the production of compounds
 CC with an antiarthritic, antiparkinsonian, neuroprotective, nootropic,
 CC immunosuppressive, cytostatic, antipsoriatic, antiinflammatory,
 CC antibacterial, antiviral, antifungal or antiparasitic activity. In
 CC addition, the disclosed sequences may be useful for gene therapy. The
 CC polypeptides or their antibodies are useful for treating many diseases
 CC such as arthritis, Parkinson's, Alzheimer's, autoimmune diseases, cancer,
 CC psoriasis, inflammatory bowel disease and infections caused by bacteria,
 CC viruses, fungi or parasites. The present sequence is that of a human
 CC polypeptide of the invention.

XX Sequence 232 AA;

XX Alignment Scores:

XX Pred. No.: 6.85e-38 Length: 232
 XX Score: 405.00 Matches: 108
 XX Percent Similarity: 56.1% Conservative: 11
 XX Best Local Similarity: 50.9% Mismatches: 52

Query Match:	9.2%	Indels:	42
DB:	5	Gaps:	5

US-10-607-806-1-G7328-T9182_COPY_7000_9500 (1-2501) x ADK36978 (1-232)

QY 600 TGGGTTTTTTTTTTTGTGTTTGTAGAGACAGAGGCTCTTGTCTGTCAACCAGGATGAGCA 659
 DB 52 TrpProPhePhe-----PhePheGlnThrGluSerArgSerValThrGlnAlaGlyVal 69
 QY 660 CAGTGTGCAACCATAGGTCATGTCAGGCTCAACCTCTCCTGAGCTCAAGGATCTGCTGAC 719
 DB 70 Gln***CysAsnLeuSerSerLeuGlnProLeuProGlyLeuLysGlnPheSerArg 89
 QY 720 CTCAGCCTCCCAAGTAGCTGGGACGAGCGTGACCCACCGCTGGCTTAATTAATAA 779
 DB 90 LeuSerLeuProSerSerTrpAspTrpArgHisMetProProArgProAlaSer---Phe 108
 QY 780 AATTTTTTTGTAGAGCTGGGCTTACTAGTTGGCCAGGCTTGTCTTAACTCTCTG-GC 838
 DB 109 ValPheLeuValGluThr-GlyPheThrMetLeuAlaArgLeuValTrpAsnSer***Pr 128
 QY 839 TTCAGCAATCTCTACCTTGGCATCCCAAGTCTGGGATTACAGGGGTGAGC----- 893
 DB 128 oSerCysAspProProIleSerAlaSerGlnAsnAlaGlyIleThrGly***SerProSe 148
 QY 894 -----CACCATGTGGGCTACTTATTTCTTTACATTCATCTTTCCAATAAATG 943
 DB 148 rLeuAlaProHisHisProArgLeu----- 156
 QY 944 TAAGATCCACAGAACAGGGATTACTGCTATTTTCTTCTTCTTTTGTGAGACAGAGTC 1003
 DB 157 -----PhePheGluThrLysSe 163
 QY 1004 TCACTTCATCACTCAACCTCCCTTCAGC-----TCACTCAACCTTCGC 1048
 DB 163 rGlySerValThrGlnAlaGlyVal-GlnTrpHisAsnProThrSerGlnProLeuP 183
 QY 1049 CTCGCGGTTCAAGYGATTCTCTGCTAAGCCTCTGAGTAGCTGGAATTACAAGCGTG 1108
 DB 183 roProGlyPheLysGlnLeuSerArgLeuSerSerSerTrpAsp***ArgTyra 203
 QY 1109 CACCACCATGCTTGGCTAATTTTGTATTTTGTAGCAGATGGGGTTTACCATGTTGC 1168
 DB 203 laProSerCysLeuAlaAsn-TyrCysIlePheSerArgAspArgAlaSerProCysTrp 222
 QY 1169 CAGGCTGGTCTCAAACTCCTGACCTCAAG 1198
 DB 223 SerGlyTrpSer***ThrProAspLeuArg 232

RESULT 11
 ADK36936
 ID ADK36936 standard; protein; 302 AA.
 AC ADK36936;
 DT 06-MAY-2004 (first entry)
 XX Novel human polypeptide SeqID9018.
 XX antiarthritic; antiparkinsonian; neuroprotective; nootropic;
 KW immunosuppressive; cytostatic; antipsoriatic; antiinflammatory;
 KW antibacterial; antiviral; antifungal; antiparasitic; gene therapy;
 KW arthritis; Parkinson's; Alzheimer's; autoimmune disease; cancer;
 KW psoriasis; inflammatory bowel disease; infection; bacteria; virus;
 KW fungus; parasite; human.
 XX Homo sapiens.
 XX Key Location/Qualifiers
 XX Misc-difference 1..302
 FT /label= OTHER
 FT /note= "OTHER= All Xaa's in this sequence are unknown
 FT amino acids or the site of a stop codon within the DNA
 FT sequence"

```

FT XX sequence"
PN XX WO200216439-A2.
PD XX 28-FEB-2002.
XX XX 05-MAR-2001; 2001WO-US004941.
PR XX 07-MAR-2000; 2000US-00519705.
XX XX 19-MAY-2000; 2000US-00574454.
XX XX (HYSE-) HYSEQ INC.
XX XX
PI XX Tang YT, Liu C, Drmanac RT;
XX XX WPI; 2002-280918/32.
XX XX
XX XX Isolated polynucleotide encoding bone marrow derived polypeptides useful
PT XX for treating, e.g., Parkinson's, Alzheimer's, cancer, arthritis, Crohn's
PT XX disease, and inflammatory bowel disease.
XX XX
XX XX Claim 20; SEQ ID NO 9018; 504pp; English.
XX XX
XX XX This invention relates to a novel isolated polynucleotide comprising a
CC XX nucleotide sequence selected from one of 1680 sequences, a mature protein
CC XX coding portion of them, an active domain of them and their complementary
CC XX sequences. The invention may be useful for the production of compounds
CC XX with an antiarthritic, antiparkinsonian, neuroprotective, nootropic,
CC XX immunosuppressive, cytostatic, antipsoriatic, antiinflammatory,
CC XX antibacterial, antiviral, antifungal or antiparasitic activity. In
CC XX addition, the disclosed sequences may be useful for gene therapy. The
CC XX polypeptides or their antibodies are useful for treating many diseases
CC XX such as arthritis, Parkinson's, Alzheimer's, autoimmune diseases, cancer,
CC XX psoriasis, inflammatory bowel disease and infections caused by bacteria,
CC XX viruses, fungi or parasites. The present sequence is that of a human
XX XX polypeptide of the invention.
XX XX
XX XX Sequence 302 AA;
XX XX
XX XX Alignment Scores:
XX XX Pred. No.: 3.02e-37 Length: 302
XX XX Score: 400.00 Matches: 110
XX XX Percent Similarity: 53.1% Conservative: 20
XX XX Best Local Similarity: 44.9% Mismatches: 72
XX XX Query Match: 9.0% Indels: 43
XX XX DB: 5 Gaps: 5
XX XX
US-10-607-806-1-G7328-T9182_COPY_7000_9500 (1-2501) x ADK36936 (1-302)
QY 597 GTTTGGGTTTCTTTCTTCTGTGT-----TTAGACAGAGGCTTGTCTCT 641
DB 81 ValCysThrPhePheIleTyrcysProPheLeuPhePheLeuArgThrGlyPheSerAla 100
QY 642 GTCACCCAGCAGTACGACAGTGTGTGCAACATAGTCTACTGAGCTCAACCTCTCTGAG 701
DB 101 LeuSerGlnAlaValGlnTrpArgAspLeuSerSerLeuGlnProProProGly 120
QY 702 CTCAGGGATCTGCTGACCTCAGCTCCAGTACGCTGAGCTGAGCTGAGCTGAGCTGAGCT 761
DB 121 LeuLysArgPheSerCysLeuSerLeuProSerSerTrpAspHisArgHisValProPro 140
QY 762 CGCTGCTGCTAATAAAAAATTTTGTGTAGACTGGGTCTTACTAGTGGCTGAGCTGCT 821
DB 141 HisProAlaAsn-PheCysIleLeuMet***ArgTrpGlyPheThrPheAlaArgMe 160
QY 822 TGTCTTAACTCTGGCTTCAAGCAATCTCTTACCTTGGCATCCCAAGTGTCTGGGATT 881
DB 160 tThrCysAsp-----LeuLeu--ThrSerAlaSerGlnThrValArgIle 174
QY 882 ACAGGGGTGAGCCACCATGTCGGCTACTTATTTCTTTTACATTCCTTCCATGTTCCATAGAA 941
DB 175 ThrGlyValSerHisHisThrArgLeu----- 183

```

```

QY 942 TGTAGATCCACAGACAGGGATTACTGCTATTTT-----CTTCCTTTC 986
DB 184 -----lIeTyr***TyPhe***Ala***SerAspIlePhePhe 196
QY 987 TTTTGTGAGACAGTCTCTACTTCATCATCACTCAACCTCCCTTCAGC----- 1032
DB 197 PhePheGluLysGluSerCysValGlyGlnAlaGlyVal-GlnTrpHisAspLeuG1 216
QY 1033 -TCACTGCAACCTCTGCTCCCGGTTCAAGYATCTCTGCTGCTAGCTCTCTGAGTAG 1091
DB 216 yserLeuGlnProProProSerGlyLeuLysGlnPheSerCysLeuAsnLeuProSerSe 236
QY 1092 CTGGAATTACACAGCTGCACCACTGCTTGGCTAATTTTCTTATTTTAGCAGAGATG 1151
DB 236 rTrpAspHisLysArgThrProHisProGlySer-PheCysIlePheSerArgAspA 256
QY 1152 GGGTTTACATGTTGCCAGGCTGGTCTCAAACTCTGACCTCAAGTATGCTGCTGCC 1211
DB 256 rgValSerProTrpGlyProGlyTrpPhe***ThrProAspLeuLysArgSerAlaLeuL 276
QY 1212 TCAGTCTCCCAAGTCTGGAATTATAGGCGTGAGTCACTGTGCTGCTGCCGATTACTGTC 1271
DB 276 euGlyLeuProLysCysAspTyArgArgGluProProArgProAlaGluValThrI 296
QY 1272 TATTT 1276
DB 296 lePhe 297
XX XX
XX XX RESULT 12
XX XX ABG68738
XX XX ID ABG68738 standard; protein; 361 AA.
XX XX AC ABG68738;
XX XX XX 07-OCT-2002 (first entry)
XX XX DE Human prostate specific protein DEX0293_107.
XX XX KW Human; prostate specific nucleic acid; PSNA; prostate cancer; PSP;
XX KW prostate specific protein; cytostatic; non-cancerous prostate disease;
XX KW gene therapy; cancer; immunostimulant; vaccine.
XX OS Homo sapiens.
XX XX WQ200255735-A2.
XX XX PD 18-JUL-2002.
XX XX PF 27-NOV-2001; 2001WO-US044363.
XX XX PR 27-NOV-2000; 2000US-0253176P.
XX XX PA (DIAD-) DIADEXUS INC.
XX XX PI Salceda S, Macina RA, Recipon H, Cafferkey R, Ali S, Sun Y;
XX XX PI Liu C, Chen S;
XX XX DR WPI; 2002-557831/59.
XX XX DR N-PSDB; ABK97631.
XX XX PT New prostate specific genes, useful for treating or diagnosing cancer, or
XX XX PT useful as vaccines for treating cancer, particularly prostate cancer, in
XX XX PT a patient.
XX XX PS Claim 11; Page 207-208; 212pp; English.
XX XX
XX XX The invention relates to a new isolated prostate-specific nucleic acid
XX XX (PSNA) molecule comprising the cDNA sequences appearing as ABK97574-
XX XX ABK97642 which encode prostate specific proteins appearing as ABG68701-
XX XX ABG68746, or a sequence hybridising to a PSNA or which has 60% sequence
XX XX homology with a PSNA. Also included are a method of determining the
XX XX presence of a PSNA in a sample, a vector comprising the PSNA, a host cell
XX XX comprising the vector, producing the polypeptide encoded by the PSNA, a

```

CC method of determining the presence of a PSP in a sample, diagnosing and
 CC monitoring the presence and metastases of prostate cancer in a patient, a
 CC kit for detecting a risk of cancer or presence of cancer in a patient, a
 CC (the kit comprising a means for determining the presence of the PSA or
 CC PSP in a sample of a patient) and a vaccine comprising the polypeptide or
 CC the nucleic acid encoding the polypeptide. The PSA, PSP and anti-PSA
 CC antibody are useful for diagnosing and treating cancer in a patient (e.g.
 CC by gene therapy). The nucleic acid molecule and polypeptide are also
 CC useful as vaccines for treating cancer, particularly prostate cancer and
 CC non-cancerous prostate diseases. The present sequence is a PSP of the
 CC invention
 XX
 SQ Sequence 361 AA;

Alignment Scores:
 Pred. No.: 3 27e-37 Length: 361
 Score: 400.00 Matches: 105
 Percent Similarity: 55.8% Conservative: 11
 Best Local Similarity: 50.5% Mismatches: 55
 Query Match: 9.0% Indels: 38
 DB: 5 Gaps: 5

US-10-607-806-1-G7328-T9182_COPY_7000_9500 (1-2501) x ABG68738 (1-361)

Qy 639 TCTGTACCCAGCATGACGACAGTGTGCAACATAGGTCACTGACGCTCAACCTCCT 698
 Db 172 SerValAlaGlnAlaGlyValGlnTrpCysAsnLeuGlySerLeuGlnProLeuProPro 191
 Qy 699 GAGCTCAAGGATCTGCTGACCTCAGCTCCCAAGTAGCTGGGACTACGAGCGTGACCA 758
 Db 192 ArgPheArgArgPheSerCysLeuSerLeuSerSerTrpAspHisArgHisAlaPro 211
 Qy 759 CCAGCTGGCTAATAAATA-----AATTTTTGTAGAGACTGGTCTTACTAGTT 812
 Db 212 ProCysLeuAlaAsnPheLeuPhePhePhePheValAlaGln-SerPheTrMetLe 231
 Qy 813 GGCCAGGCTCTCTTAACCTCTG-GCTTCAAGCAATCTCTACCTGGCATCCCAAG 871
 Db 231 uAlaArgLeuValLeuAsnSerAlaProSerGlyAspLeuProAlaSerGlnSe 251
 Qy 872 TGCTGGGATTACAGGGGTGAGCCAC----- 896
 Db 251 rAlaGlyIleThrGlyValArgHisCysThrTrpProLysSerPheLysPheAlaAspSe 271
 Qy 897 -CATGTGGGTACTATTCTTTTACATTCATCTTCCATAGAAATGTAAGATCCACAG 955
 Db 271 rHisIleGlyLeu-----AlaPheHisPheAlaPhe----- 281
 Qy 956 AACAGGGATTACTGCTATTCTTCTCTTCTTTTGGAGACAGAGTCTCATCTCATCAC 1015
 Db 282 -----PhePhePhePhePhePhePhePhePhePhePhePhePhePhePhePheAl 296
 Qy 1016 CTCACCTCCGTTCCAGC-----TCACTGCAACCTCTGCTCCCGGGTTCA 1060
 Db 296 aGlnAlaGlyVal-GlnTrpArgAspLeuGlySerLeuGlnProProProGlyPheL 316
 Qy 1061 AGYGATTCTCTGCTAAGCTCTCTGAGTACTGGAATTAACAGCTGCACCACTGCT 1120
 Db 316 ySgInPheLeuCysLeuSerLeuProGlySerTrpAspTyrArgArgAlaProProArg 336
 Qy 1121 TGGCTAATTTTTTGTATTTTATAGCAGATGGGTTTTTACCATCTGTGCCAGGCTGTCT 1180
 Db 336 lAlaAsn-PheCysIlePheSerArgAspGlyValSerProCysTrpThrGlyTrpSer 355
 Qy 1181 CAAACTCTGACCTCAAG 1198
 Db 356 GlnThrProAspLeuArg 361
 RESULT 13
 ID ABJ01057 standard; protein; 449 AA.
 XX AC ABJ01057;
 XX

XX
 DT 28-NOV-2002 (first entry)
 XX
 DE Ovary cell-specific amino acid sequence 3.
 XX
 KW Ovary cell; neoplastic ovary cell; ovary specific nucleic acid;
 KW ovary specific protein; ovarian cancer; breast cancer; vaccine;
 KW gene therapy.
 XX
 OS Homo sapiens.
 XX
 PN WO200238606-A2.
 XX
 PD 16-MAY-2002.
 XX
 PF 07-NOV-2001; 2001WO-US046459.
 XX
 PR 08-NOV-2000; 2000US-0246640P.
 XX
 PA (DIAD-) DIADEXUS INC.
 XX
 PI Sun Y, Recipon H, Salceda S, Liu C;
 XX
 DR WPI; 2002-519297/55.
 XX
 PT Polypeptide and polynucleotides present in normal and neoplastic ovary
 PT cells, useful for identifying, monitoring, staging, diagnosing,
 PT preventing and treating ovarian cancer, and non-cancerous disease states
 PT in the ovary.
 XX
 PS Claim 11; Page 207-209; 247pp; English.
 XX
 CC The invention comprises amino acid and DNA sequences which are present in
 CC normal and neoplastic ovary cells. The DNA and protein sequences of the
 CC invention are useful for determining the presence of an ovary specific
 CC nucleic acid or an ovary specific protein in a sample. The DNA and
 CC protein sequences of the invention are useful for diagnosing and
 CC monitoring the presence and metastasis of ovarian cancer and breast
 CC cancer. Amino acids ABJ01055 - ABJ01155 represent the ovary cell specific
 CC protein sequences of the invention
 XX
 SQ Sequence 449 AA;
 Alignment Scores:
 Pred. No.: 3 61e-37 Length: 449
 Score: 400.00 Matches: 103
 Percent Similarity: 58.1% Conservative: 26
 Best Local Similarity: 46.4% Mismatches: 62
 Query Match: 9.0% Indels: 31
 DB: 5 Gaps: 7
 US-10-607-806-1-G7328-T9182_COPY_7000_9500 (1-2501) x ABJ01057 (1-449)
 Qy 588 GTGGTGACAGTCTGGGTTTTTTTTTTGTTGTTAGACAGAGGTTCTTCTGTGACC 647
 Db 243 LeuIleAsnAlaLeuIleTyrPhe-----LeuArgGlnSerLeuArgSerValAla 259
 Qy 648 CAGCGATGACACAGTGGTGCACACCATAGG-----TCAGTGAGCTCAACCTCT 698
 Db 260 GlnAlaGlyValGlnTrpCysSerGlyAlaAspLeuGlySerLeuGlnProLeuProPro 279
 Qy 699 GAGCTCAAGGATCTGCTGACCTCAGCTCCAGTAGCTGGGACTACGAGCGTGACCA 758
 Db 280 GlyPheLysAlaPheProCysLeuSerLeuSerLeuSerTrpAspTyrArgSerLeuPro 299
 Qy 759 CCAGCGCTGGCTAATTAATAAATAATTTTTTTGTAGAGCTGGGTTCTTACTAGTTGGCCAG 818
 Db 300 ProCysProAlaAsnPhePheValPheLeuIleGlnThrGlyPheHisHisIleSerGln 319
 Qy 819 GCTTGTCTTAACTCTGGCTTCAAGCAATCTCTTACCTTGGCATCCCAAGCTGGG 878
 Db 320 -----IleSerIleSerAlaProCysAspProProAlaSerAlaSerGlnSerAlaGly 337

QY 1230 GGAATTATAGCGTGAAGTCACTGGCTGGCCG 1262
 Db 280 GlyIleThrGlyMetSerTyrArgAlaIrrPro 290
 RESULT 15
 AAB01399
 ID AAB01399 standard; protein; 375 AA.
 AC AAB01399;
 XX
 XX 20-OCT-2000 (first entry)
 XX
 XX Neuron-associated protein.
 XX
 XX Neuron associated protein; NEUAP; neurological disorder; epilepsy;
 KW ischemic cerebrovascular disease; stroke; cerebral neoplasm;
 KW Alzheimer's disease; Pick's disease; Huntington's disease; dementia;
 KW Parkinson's disease; demyelinating disease; meningitis; prion disease;
 KW kuru; Creutzfeldt-Jakob disease; neurofibromatosis; cerebral palsy;
 KW muscular dystrophy; central nervous system; CNS;
 KW peripheral nervous system; PNS; myopathy; schizophrenia;
 KW actinic keratosis; arteriosclerosis; atherosclerosis; bursitis;
 KW cirrhosis; hepatitis; mixed connective tissue disease; MCTD;
 KW myelofibrosis; paroxysmal nocturnal haemoglobinuria; cancer;
 KW autoimmune disease; inflammation; acquired immunodeficiency syndrome;
 KW AIDS; Addison's disease; adult respiratory distress syndrome; allergy;
 KW ankylosing spondylitis; amyloidosis; anaemia; asthma;
 KW Werner syndrome; trauma; human.
 XX
 XX Homo sapiens.
 OS
 XX
 XX W0200034477-A2.
 PN
 XX
 XX 15-JUN-2000.
 PD
 XX
 XX 10-DEC-1999; 99WO-US030408.
 PF
 XX
 XX 11-DEC-1998; 98US-00210083.
 PR
 XX 09-FEB-1999; 99US-0119365P.
 PR
 XX 16-MAR-1999; 99US-0124687P.
 XX
 XX (INCY-) INCYTE PHARM INC.
 PA
 XX
 XX Tang YT, Yue H, Baughn MR, Hillman JL, Lal P, Au-Young J;
 PI Yang J, Lu DAM, Azimzai Y;
 FI
 XX WPI; 2000-423423/36.
 DR
 XX
 XX New human neuron-associated proteins and polynucleotides encoding them,
 PT useful for diagnosis, treatment and prevention of cell proliferative
 PT disorders including cancer, neuronal and neurological disorders.
 XX
 XX Disclosure; Page 143-144; 145pp; English.
 PS
 XX
 XX Human neuron-associated proteins (NEUAP) can be used for treating or
 CC preventing a disorder associated with decreased expression or activity of
 CC NEUAP. Antagonists of NEUAP are useful for treating or preventing
 CC disorder associated with increased expression or activity of NEUAP. NEUAP
 CC or their fragments or derivatives are useful for treating neurological
 CC disorder such as epilepsy, ischemic cerebrovascular disease, stroke,
 CC cerebral neoplasms, Alzheimer's disease, Pick's disease, Huntington's
 CC disease, dementia and Parkinson's disease. NEUAPs are also useful for
 CC treating other demyelinating diseases, bacterial and viral meningitis,
 CC prion diseases including kuru, Creutzfeldt-Jakob disease, nutritional and
 CC metabolic diseases of the nervous system, neurofibromatosis, other
 CC developmental disorders of the central nervous system, cerebral palsy,
 CC neuroskeletal disorders, autonomic nervous system disorders, cranial
 CC nerve disorders, spinal cord diseases, muscular dystrophy and other
 CC neuromuscular disorders, peripheral nervous system disorders, inherited,
 CC metabolic, endocrine, and toxic myopathies, mental disorders including
 CC mood, anxiety and schizophrenic disorders, a cell proliferative disorder
 CC such as actinic keratosis, arteriosclerosis, atherosclerosis, bursitis,
 CC cirrhosis, hepatitis, mixed connective tissue disease (MCTD),

CC myelofibrosis, paroxysmal nocturnal haemoglobinuria, cancers of the
 CC adrenal gland, bladder, bone, bone marrow, brain, breast, cervix, and an
 CC autoimmune/inflammatory disorder such as acquired immunodeficiency
 CC syndrome (AIDS), Addison's disease, adult respiratory distress syndrome,
 CC allergies, ankylosing spondylitis, amyloidosis, anemia, asthma, Werner
 CC syndrome, complications of cancer, hemodialysis, and extracorporeal
 CC circulation, viral, bacterial, fungal parasitic, protozoal, and
 CC helminthic infections, and trauma. This protein was designated g3002527
 XX
 XX Sequence 375 AA;
 SQ
 Alignment Scores:
 Pred. No.: 6.58e-37 Length: 375
 Score: 397.50 Matches: 105
 Percent Similarity: 56.0% Conservative: 12
 Best Local Similarity: 50.2% Mismatches: 60
 Query Match: 9.0% Indels: 32
 DB: 3 Gaps: 4
 US-10-607-806-1-G7328-T9182_COPY_7000_9500 (1-2501) x AAB01399 (1-375)
 QY 597 GTTTGGGTTTTTTTGGTTGTTTGTAGACAGAGGTCTGTCTCACCAGGATGA 656
 Db 194 IlePheIlePheIlePheAenPheLeuArgGlnSerLeuAsnSerValThrGlnAlaGly 213
 QY 657 GCACAGTGGTGCACCAACATAGGTCTACCTGACGCTCAACCTCTGAGCTCAAGGGATCTGCT 716
 Db 214 ValGlnTrpArgAsnLeuGlySerLeuGlnProLeuProGlyPheLeuLeuPheSer 233
 QY 717 GACCTCAGCTCCCAAGTAGCTGGGACTAGAGCGTGACACCCAGCGCTGGCTAAATTA 776
 Db 234 CysProSerLeuLeuSerSerTrpAspTyrArgArg---ProProArgLeuAlaAenPhe 252
 QY 777 AAAAAATTTTGTAGAGACTGGTCTTACTAGCTGGCGAGGCTGTCTTAACCTCTG 836
 Db 253 PheValPheLeuValGlu-MetGlyPheThrMetPheAlaArgLeuLeuLeuSer-- 271
 QY 837 GCTTCAAGCAATCTCTCTACCTTGGCATCCCAAGTGGGATTACAGGGGTGAGCCAC 896
 Db 272 GlyProCysAspLeuProAlaSerAlaSerGlnSerAlaGlyIleThrGlyValSerHis 291
 QY 897 CATGTGCGCTACTTATTTCTTACATTCCTATTCATCTTTTCCAATAGATGTAAGATCCACA 956
 Db 292 HisAlaArgLeuIlePheAenPhe-----Cys----- 300
 QY 957 ACAGGGATTTACTGCTATTTCTTCTTCTTTTCTTTTGTAGACAGAGTCTCACTCATCACC 1016
 Db 301 -----LeuPheGluMetGluSerHisSerValThr 310
 QY 1017 TCAA-----CCTCGTTCAGCTCACTGCAACCTCTGTCTCCCGGGTTCAA 1061
 Db 311 -GlnAlaGlyValGlnTrpProAsnLeuGlySerLeuGlnProLeuProGlyLeuLeu 330
 QY 1062 GYGATTTCTCTGCTAAGCTCTCTGAGTAGCTGGAATTACAAGCGTGACACCATGCTT 1121
 Db 330 sArgPheSerCysLeuSerLeuProSerSerTrpAspTyrGlyHisLeuProHisPr 350
 QY 1122 GGCTAATTTTGTATTTTGTAGCAGAGATGGGCTTTTACCATGTTGCCAGGCTGCTC 1181
 Db 350 oAlaAsn-PheCysIlePheIleArgGlyValSerProTyrLeuSerGlyTrpSerG 370
 QY 1182 AAATCTCTGACCTCAAG 1198
 Db 370 InthrProAspLeuArg 375

Search completed: February 8, 2006, 10:43:53
 Job time : 422.846 secs

GenCore version 5.1.7
Copyright (c) 1993 - 2006 Bioceleration Ltd.

OM nucleic - protein search, using frame_plus_n2p model

Run on: February 8, 2006, 10:44:07; Search time 17.0675 Seconds
(without alignments)
2819.838 Million cell updates/sec

Title: US-10-607-806-1-G7328-T9182_COPY_7000_9500
Perfect score: 4426
Sequence: 1 gtcgtgctactgtgtccag.....tttgagaccagcctggacaa 2501

Scoring table: BLOSUM62
Xgapop 10.0, Xgapext 0.5
Ygapop 10.0, Ygapext 0.5
Fgapop 6.0, Fgapext 7.0
Delop 6.0, Delext 7.0

Searched: 283416 seqs, 96216763 residues

Total number of hits satisfying chosen parameters: 566832

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Command line parameters:

-MODEL=frame+n2p.model -DEV=xlp
-Q=/abs/ABSFEB_spool/WARTIN10607806/runat_08022006.092020_27698/app_query.fasta_1
-DBS=PIR -QFMT=fastan -SUFFIX=n2p.rpr -MINMATCH=0.1 -LOOPCL=0 -LOOPEXT=0
-UNITS=bits -START=1 -END=1 -MATRIX=blosum62 -TRANS=human40.cdi -LIST=45
-DOCALIGN=200 -THR_SCORE=pct -THR_MAX=100 -THR_MIN=0 -ALIGN=15 -MODE=LOCAL
-OUTFMT=ptc -NORM=ext -HEADSIZE=500 -MINLEN=0 -MAXLEN=2000000000 -HOST=abs02p
-USER=WARTIN10607806 @CCN 1 1 92 @runat_08022006.092020_27698 -NCPU=6 -ICPU=3
-NO_MMAP -NEG_SCORES=0 -WAIT -DSPBLOCK=100 -LONGLOG -DEV_TIMEOUT=120
-WARN_TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5 -FGAPOP=6 -FGAPEXT=7
-YGAPOP=10 -YGAPEXT=0.5 -DBLOP=6 -DELEXT=7

Database : PIR_80:*
1: pir1:*
2: pir2:*
3: pir3:*
4: pir4:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	416	9.4	673	4 F40201	artifact-warning s
C 2	402.5	9.1	613	4 C40201	artifact-warning s
C 3	391.5	8.8	673	4 F40201	artifact-warning s
C 4	391	8.8	613	4 C40201	artifact-warning s
C 5	338	7.6	627	4 A40201	artifact-warning s
C 6	332.5	7.5	627	4 A40201	artifact-warning s
C 7	276.5	6.2	597	4 E40201	artifact-warning s
C 8	269.5	6.1	418	2 S41044	chromosomal protei
C 9	266.5	6.0	597	4 E40201	artifact-warning s
10	264	6.0	196	2 I38022	hypothetical prote
11	258	5.8	148	1 PSUO	phospholipase A2 (
C 12	248.5	5.6	301	4 B40201	artifact-warning s
C 13	239.5	5.4	301	4 B40201	artifact-warning s
14	239	5.4	579	4 D40201	artifact-warning s

C 15	232.5	5.3	100	2 A46010	X-linked retinopat
C 16	223.5	5.0	579	4 D40201	artifact-warning s
C 17	218	4.9	146	1 PSDG	phospholipase A2 (
18	216	4.9	144	2 JN0480	phospholipase A2 (
19	216	4.9	146	1 PSRT	phospholipase A2 (
C 20	214.5	4.8	499	2 S65657	alpha-1C-adrenergi
C 21	209	4.7	145	1 PSBOA	phospholipase A2 (
C 22	206.5	4.7	418	2 S41044	chromosomal protei
C 23	206	4.7	196	2 I38022	hypothetical prote
C 24	205	4.6	146	1 PSPGA	phospholipase A2 (
25	202	4.6	124	1 PSPGA2	phospholipase A2 (
26	201.5	4.6	132	1 PSBOA	phospholipase A2 (
27	188	4.2	146	1 S34049	phospholipase A2 (
C 28	172	3.9	46	2 I54375	gene NF2 protein -
C 29	162	3.7	39	2 I54374	gene NF2 protein -
C 30	160	3.6	369	2 A53959	thromboxane A-2 re
31	159	3.6	119	2 S00227	phospholipase A2 (
32	157	3.5	125	2 S65624	phospholipase A2 (
C 33	156	3.5	53	2 A42442	integrin beta-1 ch
C 34	155.5	3.5	407	2 T02670	probable thromboxa
C 35	154.5	3.5	79	2 A56194	thromboxane A-2 re
C 36	149.5	3.4	119	1 PSN33K	phospholipase A2 (
C 37	147	3.3	841	1 I78885	serine/threonine-s
38	144.5	3.3	146	1 PSNUAF	phospholipase A2 (
39	144.5	3.3	499	2 S65657	alpha-1C-adrenergi
40	143.5	3.2	119	1 PSN22K	phospholipase A2 (
C 41	143.5	3.2	124	2 A47582	B-cell growth fact
42	143	3.2	133	1 PSOXG	phospholipase A2 (
43	142.5	3.2	100	2 A46010	X-linked retinopat
44	140.5	3.2	119	1 PSRIA	phospholipase A2 (
45	139.5	3.2	118	2 S01801	phospholipase A2 (

ALIGNMENTS

RESULT 1

F40201
artifact-warning sequence (translated ALU class F) - human
C;Species: Homo sapiens (man)
C;Date: 31-Mar-1992 #sequence_revision 11-Aug-1995 #text_change 19-May-2000
C;Accession: F40201
R;Claverie, J.M.
personal communication, 1992
A;Reference number: A40201
A;Accession: F40201
A;Molecule type: DNA
A;Residues: 1-673 <Cl>
A;Cross-references: UNIPARC:UPI000017CECD
R;Claverie, J.M.
Genomics 12, 838-841, 1992
A;Title: Identifying, coding exons by similarity search: Alu-derived and other potential
A;Reference number: A40200; MUID:92241891; PMID:1572661
A;Comments: annotation
C;Comment: This "warning" entry is a conceptual translation in all 6 reading frames of
in-frame stop codons are shown as 'x'.
C;Comment: Any significant similarity of a predicted protein sequence to a portion of t

Alignment Scores:			
Pred. No.:	3,38e-35	Length:	673
Score:	416.00	Matches:	131
Percent Similarity:	41.8%	Conservative:	17
Best Local Similarity:	37.0%	Mismatches:	85
Query Match:	9.4%	Indels:	122
DB:	4	Gaps:	9

US-10-607-806-1-G7328-T9182_COPY_7000_9500 (1-2501) x F40201 (1-673)

QY	532	AGAAAACTGAGGCTCAGAGGCTGATTTGCCACAGATCACACACCTGTAGTGG	591
DB	328	LySLysLysAsnProGlnlys-----Gln*****PheCysPheTrp	342
QY	592	TGACAGTTTGGGTTTTTTTTTTTGTGTGTTTGA-----GAGACAGGGCTTGTCTGTC	644

Db	343	GlyPhePhePhePhePheCysPheValPheGluLeuGlyProCysSerAsp	362
Qy	645	ACC CAG CAG CAG CAG CAG TGG TGC CAA CCA TTAG GTC TACT G CAG C C T C A A C C T C C T G A G C T C	704
Db	363	ThrProAlaGlyValGlnTrp-GlnValLeuAlaHisCysSerLeuAsnLeuLeuGlySer	382
Qy	705	AAGGGATCTGCTGACCTCAGCCTCCCAAGTAGCTGGGACTACGAGCCTGCACACCCACACGC	764
Db	382	rSerAspSerProAlaSerValSerArgValAlaGlyIleThrGlyMetGlyArgHis	402
Qy	765	CTGGCTAATATAAAAAATTTTTTGTAGACACTGGGTCTTACTACGTTGGCCAGGCTGT	824
Db	402	rTrpLeuIleTyrVal--PheLeuIleGlnThrGlnPheHisValAspGlnAlaGly	421
Qy	825	CTTAAACTCTCGGTTCACGAATCTCTACCTTGGCATCCCAAAAGTGTGGGATTACA	884
Db	422	LeuLysLeuLeuThrSerSerAspLeuProSer-TipSerProLysValLeuGly**G	441
Qy	885	GGGTGAGCCACC-----ATGTGGCGTACTATTATTCCTTTA	920
Db	441	nAla***AlaThrThrProSer*****PheValPheGlyGlyPhePhePhe	461
Qy	921	CATTCCA-----	927
Db	461	ePheAlaLeuPheLeuPheLeuArg***AlaLeuAlaLeuThrProArgLeuGluCys	481
Qy	928	-----TCCTTCCAAATAGAAATGTAGATCCACA-----GAACAGGGATTACTGCCTAT	974
Db	481	rGlyLysPheTrpLeuThrAlaAlaSerThrSerTrpValGlnAlaIleLeuLeuProLe	501
Qy	975	T-----TTCCTTCCTTC	986
Db	501	uSerProVal**LeuGlyLeuGlnAlaTrpAlaAlaIleProGly***PheMetTyrPh	521
Qy	987	TTTTTTTGAGACAGAGTCTCACT-----TCATCACTCAAA--	1020
Db	521	e*****ArgHisSerPheThrMetLeuIleArgLeuValSerAsnSer***ProGlnVa	541
Qy	1021	-----CCTCCG-----	1026
Db	541	lIleCysProGlyLeuProLysCysTrpAspArgArgGluProProHisProAl	561
Qy	1026	-----	1026
Db	561	a*****LeuPheLeuGlyValPhePhePhePheLeuLeuCysPheCysPhe**	581
Qy	1027	-----TTCAGCTCACTGCA	1040
Db	581	*AspArgProLeuLeu***HisProGlyTrpSerAlaValAlaSerPheGlySerLeuGl	601
Qy	1041	ACCTCTGCCTCCGGGTTCAGYGATTCTCTCTGCTAAGCCTCCTGAGTAGCTGAATTA	1100
Db	601	nProGlnProProGlyPheLysArgPheSerCysLeuProCysSerTrpAspTyr	621
Qy	1101	CAAGCGTGACCAACCATGCTTGGCTAAATTTTTTGTATTTTTPAGCAGAGATGGGGTTTTAC	1160
Db	621	rArgHisGlyProProPheLeuAlaAsn-LeuCysIlePheAsnArgAspThrValSerP	641
Qy	1161	CATCTGCCAGGTGGTCTCAACTCTGACCTCAAGTGATCGCTGCTCGCTCGATCTCC	1220
Db	641	roCys***SerGlyTrpSerGlnThrProAspLeuLys***SerAlaLeuLeu-ValSer	660
Qy	1221	CAAAAGTGTGGAAATTTATAGGCGTGAGTCAC	1250
Db	661	GlnSerAlaGlyMetThrGlyValSerHis	670

R;Claverie, J.M.
personal communication, 1992
A;Reference number: A40201
A;Accession: C40201
A;Molecule type: DNA
A;Residues: 1-613 <CLA>
A;Cross-references: UNIPARC:UPI000017CECA
R;Claverie, J.M.
Genomics 12, 838-841, 1992
A;Title: Identifying coding exons by similarity search: Alu-derived and other potentially
A;Reference number: A40200; MUID:922441891; PMID:1572661
C;Contents: annotation
C;Comment: This "warning" entry is a conceptual translation in all 6 reading frames of on
in-frame stop codons are shown as 'X'.
C;Comment: Any significant similarity of a predicted protein sequence to a portion of the

Alignment Scores:					
Pred. No.:	9,08e-34	Length:	613		
Score:	402.50	Matches:	162		
Percent Similarity:	40.9%	Conservative:	48		
Best Local Similarity:	31.6%	Mismatches:	151		
Query Match:	9.1%	Indels:	152		
DB:	4	Gaps:	17		
US-10-607-806-1-GT328-T9182_COPY_7000_9500 (1-2501) x C40201 (1-613)					
Qy	1281	AAGAAATAATACAGTAATCGGCAGGACACAGTCAGTCAGCCCTATAATTCCAGCACTTT	1222		
Db	97	LysLysLysLys*****AlaGlyCysSerGlySerArgLeu***SerGlnHisPhe	116		
Qy	1221	GGGAGACTGAGCGGCGAGATCACTTAGCTCAGGTACGAGTTTTGAGACCAGCTGGCACAT	1162		
Db	117	GlyArgGlnArgTrpValAspHis--GluAlaArgSerSerA+gProAlaTrpProThrT	136		
Qy	1161	GCTAAAACCCCATCTCTGCTAAAAATACAAAAAATAGCCAACGATGGTGTCACGCCTT	1102		
Db	136	rPGLnAnProlIseSerThrLYeAsnThrLysAsn***LeuGlyMetVal-ArgAlaPro	155		
Qy	1101	GTAATTCAGCTACTCAGGAGGCTTAGCGAGGAGAATCRCTTGAACCCGGGAGGCACAGG	1042		
Db	156	ValValProArgThrGlnLysAlaGluAlaGlyGluTrpHisLysProGlyA+gTrpSer	175		
Qy	1041	TTCAGTCAGCTGAA-----CGGAGGTTCAGGTGCATCAAGTCGAGACTCTG	997		
Db	176	LeuGln***AlaGluIleSerProLeuHisSerSerLeu-GlyAspargValargLeuCy	195		
Qy	996	TCTCAAAAAAGAAAGAAATAGCGACGTAATCCCTGTTCTGTGGATCTTACATTCTA	937		
Db	195	sLeuLysLysLysLysLys-----	202		
Qy	936	TTGAAAGATGGTAATGTAAAGAAATAAGTAGCCGACATGCTGGCTCACCCCTGTAATCC	877		
Db	203	-----lys*****ArgGlyAlaValAlaHisAlaCyAsnPr	216		
Qy	876	CAGCACTTTGGGATGCCAGGTAGGAGATTTGCTTGAAGCCAGGAGTTTAAGACAAGCCT	817		
Db	216	oserThrLeuGlyLysGly-GlyTrpile-MetArgProGlylValargAspGlnPro	235		
Qy	816	GGCCAACGTAGTACCCACTCTCTACAAAAAATTTTTTAATTAGCCAGGCGTGGT	757		
Db	236	GlyGlnHisGlyLysThrProPheLeu-LeuLyslleGlnLysIleSerTrpAlaTrpCy	255		
Qy	756	GTGCACGCTCGTAGTCCCAGCTACTTGGGAGGCTGAGGTGACGACATCCCTTGAGCTCAG	697		
Db	255	s-GlyArgLeu***SerHisValArgArgArgLeuArgGlnGlnAsnGlyIleAsnProG	275		
Qy	696	GAGGTTGAGGCTGCAGTACCTATGTTGCACCACTGCTCATGCTCGGTGGTGCACAGC	637		
Db	275	lyGlyGlyAlaCyserGluProArgSerArgHisCysThrProAlaTrpValThrGlu*	295		
Qy	636	AAGACCCCTGCTCTTAAACAAACAAAAA-----	605		

C;Species: Homo sapiens (man)
C;Date: 31-Mar-1992 #sequence_revision 11-Aug-1995 #text_change 19-May-2000
R;Claverie, J.M.
personal communication, 1992
A;Reference number: A40201
A;Accession: C40201
A;Molecule type: DNA
A;Residues: 1-613 <CLA>
R;Cross-references: UNIPARC:UPI000017CECA
R;Claverie, J.M.
Genomics 12, 838-841, 1992
A;Title: Identifying coding exons by similarity search: Alu-derived and other potentially
A;Reference number: A40200; MUID:92241891; PMID:1572661
C;Contents: annotation
C;Comment: This "warning" entry is a conceptual translation in all 6 reading frames of c
in-frame stop codons are shown as 'X'.
C;Comment: Any significant similarity of a predicted protein sequence to a portion of th

Alignment Scores:
Pred. No.: 1.51e-32 Length: 613
Score: 391.00 Matches: 153
Percent Similarity: 40.2% Conservative: 25
Best Local Similarity: 34.5% Mismatches: 131
Query Match: 8.8% Indels: 134
DB: 4 Gaps: 14

US-10-607-806-1-G7328-T9182_COPY_7000_9500 (1-2501) x C40201 (1-613)

QY 2 TCGTGTCAGCTGCTCCAGCTGGTAACAGAGCAACTGCTCTCAAAAAAATAATGC 61
DB 284 SerArgHisCysThrProAlaIrrpValThrGlu***AspSerValSerLysLysLys--- 302
QY 62 TTTCATAAATATATGATGATAAAGGACTTATATTTT----- 97
DB 303 -----LysLysLys*****PhePhePhePhePheGluThr 317
QY 98 TCAAGCCATAGGATCATTTCTCTGAGAGCATCTGGCGAGTATCCCGACCTGTTCTCTG 157
DB 318 GluSerHisSerValThrGlnAlaGlyValGlnTrpArgAsp----- 331
QY 158 AGAGTGGCGAGGTGAGGGCTGACCTATTGCTCTGCACCTTACTCTCTATCTCAGCTGTCCCT 217
DB 332 -----LeuGlySerLeuGlnAlaProPro 339
QY 218 CCCACTTTCCAGGTGTGCCAGACATGACAACTGCTAYGACAGGCCAAGAGCTGGA 277
DB 340 ProGlyPhe-----MetPro-Phe----- 345
QY 278 CAGCTGTAATTTCTGCTGGACAMMCGTACACCCACCTATTTCATCTCGTCTCTGG 337
DB 346 -SerCysLeuSerLeuLeu-----ArgThrTrpAspTyr----- 356
QY 338 CTGGCAATCACCTGTAGCAGTAGGTTTATCCCTTCTTCACCTATGCAATCTAGTTGGT 397
DB 356 ----- 356
QY 398 TCTCAGTAGGCCGGGGGAAATAATAGTAACACAGCCATGATTAGTGTAAATTTCTTT 457
DB 357 -----ArgArgPro-----HisHisAlaGlnLeuLeuPheCys 367
QY 458 GGTTCTGGCGAGTCTCTCTTTAATCTT-----CAGAACACACTATGGGATA 505
DB 367 sillePheSerArgAsnGlyValLeuProCysTrpProGlyTrpSerArgThrProGly-L 387
QY 506 GGTACAAATTCCTCACTTAACAGATAAGAAACTGAGGCTCAGAGGCTGAGCTATTG 565
DB 387 euMetIleHisProLeuProProLysValLeuGlyLeuGlnAla***Ala----- 404
QY 566 CCCAAGATCACACAGCTGTGAAGTGGTGACAGTTGGTTTTTTTGTGTTGTAGTA 625
DB 405 -----ThrAlaProArgPhe-----PhePhePhePhePhe*****LeuA 419

QY 626 GACAGGCTCTGCTCTGTCAACCCAGCATGAGCACAGTGTGTCACCATAGGTCACTGCA 685
DB 419 rgGlnSerLeuThrLeuSerProArgLeuGluCysAsnGlyGluIleSerAlaHisCysL 439
QY 686 GCCTCAACCTCTCTGAGCTCAAGGGATCTGCTGACCTCAGCCTCCCAAGTAGCTGGACTA 745
DB 439 ysLeuHisLeuProGlyLeuCysHisSerProAlaSerAlaPheCysValArgGlyThr 459
QY 746 CGAGCGTCACACACAGCGCTGCTGCTTAATAAAAAATTTTGTAGAGACTGGGTCTTA 805
DB 459 hrGlyAlaArg-ThrMetProSer***PhePheVal-PheLeuValGluMetGlyPheCys 478
QY 806 CTACGTTGGCCAGGCTGTCTTAAACTCTCGCTTCAAGCAATCTCTACCTTGGCATC 865
DB 478 shisValGlyGlnAlaGlyLeuLeuLeuAlaSer***SerThrHisLeuCysLeu 497
QY 866 CCAAAGTGTGGGATTACAGGGGTGAGCCA-----CCATGTGCGGCTACTTATTTCTTT 919
DB 498 ProLysCysTrpAspTyrArgGluProLeuHisProAla*****PhePhePhe 517
QY 920 ACATTCCATCTTTCCAATAGATGTAAGATCCACAGACAGGATTTACTGCCTATTTTCT 979
DB 517 ----- 517
QY 980 TCCTTTCTTTTGTAGACAGAGTCTCAGCTTCATCAC-----CTCAACCTC 1024
DB 518 ---PhePhePhe***AspArgValSerLeuCysHisProGlyTrpSerAlaMet-AlaAr 536
QY 1025 CGTTACAGCTCACTGCAACCTCTCGCTCCCGGGTTCAGYGATTCTCTCGCTAAGCTCC 1084
DB 536 gSerArgLeuThrAlaSerSerThrSerArgValTyAlaIleLeuLeuProGlnProSe 556
QY 1085 TGAGTAGCTGGAATTAACAAGCGTGACACCATGCTGTGCTAATTTTGTATTTTAGC 1144
DB 556 rAlaTyValGlyLeuGlnAla-ProAlaProCysProAlaAsnPheLeuTyPhe*** 576
QY 1145 AGAGATGGGTTTTTACCATGTTGCCAGGCTGTCTCAAACTCTGACCTCAAGTGATCT 1204
DB 576 **LysTrpGlyPheAlaMetLeuAlaArgLeuValSerAsnSerTrpProHis--AspPr 595
QY 1205 GCTGCTCAGTCTCCCAAAGTGTGGAATATATAGCGGTGAGTCACTGT 1253
DB 595 oProThrPheAlaSerGlnSerAlaGlyIleThrGlyValSerHisCys 611

RESULT 5

A40201
artifact-warning sequence (translated ALU class A) - human
C;Species: Homo sapiens (man)
C;Date: 31-Mar-1992 #sequence_revision 11-Aug-1995 #text_change 19-May-2000
C;Accession: A40201
R;Claverie, J.M.
personal communication, 1992
A;Reference number: A40201
A;Accession: A40201
A;Molecule type: DNA
A;Residues: 1-627 <CLA>
A;Cross-references: UNIPARC:UPI000017CEC8
R;Claverie, J.M.
Genomics 12, 838-841, 1992
A;Title: Identifying coding exons by similarity search: Alu-derived and other potentially
A;Reference number: A40200; MUID:92241891; PMID:1572661
C;Contents: annotation
C;Comment: This "warning" entry is a conceptual translation in all 6 reading frames of o
in-frame stop codons are shown as 'X'.
C;Comment: Any significant similarity of a predicted protein sequence to a portion of th

Alignment Scores:
Pred. No.: 6.37e-27 Length: 627
Score: 338.00 Matches: 144
Percent Similarity: 43.0% Conservative: 38
Best Local Similarity: 34.0% Mismatches: 129

```
Query Match: 7.6% Indels: 112
DB: 4 Gaps: 12
US-10-607-806-1-G7328-T9182_COPY_7000_9500 (1-2501) x A40201 (1-627)

QY 1247 ACTCAGCCCTATTAATCCACACACTTTGGGAGACTGAGGCGAGGAGATCACTTGAGGTGAC 1188
   ::::: ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 5 SerHisAlaCysAsnProSerThrLeuGlyGlyGlnGlyArgIleThr--ArgSerG 24
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

QY 1187 GAGTTTGAGACCAAGCTGGGCAACATGTTAAACCCCATCTCTGCTAAATAACAAAAA 1128
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 24 LysAspArgAspHisProGlyTyHisGlyGlnThrProSerLeuLeuLysIleGlnLysI 44
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

QY 1127 TTAGCCCAAGCATGGTGTGCACCTTGTAAATCCAGCTACTCAGGAGGCTTGGCCAGGAG 1068
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 44 LysSerArgAla-***CysGlyArgLeu***SerGlnLeuLeuGlyArgLeuArgGlnGlu 63
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

QY 1067 AATCRCTTGNACCCGGGAGGAGGAGGTGGAGTGAG-----CTGAACGGA 1023
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 64 AsnGlyValAsnProGlyGlyGlyAlaCysSerGluProArgSerArgHisCys-ThrPr 83
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

QY 1022 GGTGTAGGTGATGAGTACACTGTCTCAAAAAGAAAGAAAGAAATAGCCAGTAAAT 963
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 83 oAlaLeuAlaThrGluArgAspSerValSerGlnLys----- 95
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

QY 962 CCCTGTTCTGTGGATCTTACATTTGTAATGGAAGATGGAATGTAAGAAATAAGTAGCCG 903
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 96 -----AsnLysAsnLysLys***Lys*****Al 106
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

QY 902 CACATGGTGGCTCACCCCTGTAAATCCAGACACTTTGGGATGCCAAGTAGGAGGATGCT 843
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 106 aGlyArgGlyLeuThrProValIleProAlaLeuThrGluAlaLysAlaGlyLys--Ser 125
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

QY 842 TGAAGCCAGAGTTTAAGACAACTGGCCAACTAGTAAGACCCAGCTCTCTACAAAAA 783
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 126 ArgGlyGlnGlnGlnGlnLeuAlaThrValLysProArgLeuTyf-***Ly 145
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

QY 782 ATTTTATTAATAGCCAGGCTGGTGTGCACCTGCTAGTCCAGCTACTTGGGAGGCT 723
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 145 sTyLysLysLeuAlaGlyArgSerAlaGly-AlaCysSerProSerTyLysGlyGly* 165
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

QY 722 GAGGTGAGCAGATCCCTGTGAGCTCAGAGGTTGAGCTGCGAGTACCTATGTTGCCACCA 663
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 165 **GlyArgArgMetAla***ThrArgGluAlaGlnLeuAlaValSerArgAspArgAlat 185
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

QY 662 CTGTGCTCATGCTGGTGACAGACGACCTGTCTCTAACAACACACAAAAA 603
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 185 hrAlaLeuGlnProTrpArgGlnSerGlnThrProSerArgLysLysIleLysThrLysA 205
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

QY 602 CCAAAA-----CTGTCAACCTTACAAAGCTGTGTGATCTTGGGCA 564
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 205 snLysLys*****ProGlyAlaValSerArgLeu***Ser----- 219
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

QY 563 AATAGCTCAGCCTTCTGAGCCTCAGTTTCTTATCTGTTAAGTGAGGATAATTGTACCTA 504
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 219 ----- 219
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

QY 503 TCCCATAGTGTGTTCTGAGGATTAAAGGAGACACTGCCAGACCAACAGAAAATTAACAC 444
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 220 -----GlnHisPheGlyArgProArgArg-----A 228
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

QY 443 TAAATCATGGCTGTTGTTACTATTATTTCCCGCCGCTACTGACAGAACCACTAGAAATTC 384
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 228 laAspHis----- 233
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

QY 383 ATAGGTCAAGGAGGAGGATAAACCCTACTGTACAGGTGATGTCGCG---AGCCAGAGACGA 327
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 233 rgArgSerArgPro-----SerTrpLeuProArg***AsnProValSerThrLysAsnT 251
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

QY 326 GTATGATAGGTGTGGGTGACGGKTTGTCCAGACAGAAATTTACAGCTGTCCAGCTTCTT 267
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 251 hrLysAsn***ProGlyValValArgAlaProVal-----ValProAlaThrT 267
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
```

```
QY 266 GGCTGCTCCTAGCAGTGTGTCTGTGGCAGCACCTGGAAAGTGGAGGAGGACAGCTG 207
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 267 tp-----Glu-AlaGluAlaGlyGlnTrpArgGluProGly 278
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

QY 206 AGATAGGAGTAAAGTCAGAGCAATAGGTGAGCCCTCACCTGCCACTCTCTCAGGAACAGGT 147
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 279 Arg-----ArgSerLeuGln***AlaGluIleAlaProLeuHisSerLeu 294
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

QY 146 GGGGATGACTTCCCAAGATGCTTCAAGGAGAATGATCTGTGCTTGAATAATATAAG 87
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 295 GlyAspArgAlaArgLeuArgLeuGlyLysLys***LysGlnLysIleLysLys***** 314
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

QY 86 TCCTTTTATCATATATTTATGAAAGCATTTTTTTTTTTTGGACACAGAG 38
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 315 ***PhePheTyR-PheLeuPhe-----LeuPhePheSerGluThrGlu 328
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

RESULT 6
A40201
artifac-warning sequence (translated ALU class A) - human
C:Species: Homo sapiens (man)
C:Date: 31-Mar-1992 #sequence_revision 11-Aug-1995 #text_change 19-May-2000
C:Accession: A40201
R:Claverie, J.M.
personal communication, 1992
A:Reference number: A40201
A:Accession: A40201
A:Molecule type: DNA
A:Residues: 1-627 <CLA>
A:Cross-references: UNIPARC:UPI000017CEC8
R:Claverie, J.M.
Genomics 12, 838-841, 1992
A:Title: Identifying coding exons by similarity search: Alu-derived and other potential
A:Reference number: A40200; MUID:92241891; PMID:1572661
A:Contents: annotation
C:Comment: This "warning" entry is a conceptual translation in all 6 reading frames of
in-frame stop codons are shown as 'X'.
C:Comment: Any significant similarity of a predicted protein sequence to a portion of the
```

Alignment Scores:

Pred. No.:	2,44e-26	Length:	627
Score:	332.50	Matches:	114
Percent Similarity:	46.6%	Conservative:	18
Best Local Similarity:	40.3%	Mismatches:	86
Query Match:	7.5%	Indels:	65
DB:	4	Gaps:	4

US-10-607-806-1-G7328-T9182_COPY_7000_9500 (1-2501) x A40201 (1-627)

```
QY 597 GTTTCGGTTTTTTTTTTGTTGTTTGTAGACAGAGGTCTGTCTGTCAACCAGGCATGA 656
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 320 LeuPheLeuPhePheSer-----GluThrGluSerArgSerValAlaLysAlaGly 336
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

QY 657 GCACAGTGGTGCACACCATAGGTCACTGCAGCCCTCAACCTCTCAGCTCAAGGATCTGCT 716
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 337 ValGlnTrpArgAspLeuGlySerLeuGlnAlaProProGlyPheThrProPheSer 356
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

QY 717 GACCTCAGCCTCCCAAGTAGCTGGGACTACGAGCGGTGCACACACCGCTCGGCTAATTAA 776
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 357 CysLeuSerLeuProSerSerTrpAspTyArgArgPro-HisTyAlaArgLeuIlePh 376
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

QY 777 AAAAAATTTTTTTGTAGAGACTGGGTCTTACTAGTGTGGCCAGCTTGTCTTAACCTCCG 836
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 376 eCysIle-PheSerArgAspGlyValSerProTrp***ProGlyTrpSerArgSerProA 396
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

QY 837 GCTTCAAGCAATCTCTTACCTTGGCATCCCAAGTGTGGGATTACAGGGGTGAGCCAC 896
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 396 spLeu--ValIleArgProProTrpProProLysValLeuGlyLeuGlnAla----- 412
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

QY 897 CATGTGGCGGTACTATTATTCTTTTACATCCCATCTTTCCAAATAGATGTAAGATCCACAGA 956
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 413 -----****AspArgAlaArg** 418
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
```

```
QY 957 ACAGGGATTACTGCTATTCTTCTCTTTTGGAGACAGAGTCTCATCTTCATCACC 1016
Db 418 *****PhePheIlePheCysPheTyrPhePheProArgArgSerLeuAlaLeuSerPr 438
QY 1017 TCACCTCCGT-----TCAGCTCACTGCAACCTCTGCTCCCGGGTTCAA 1061
Db 438 oArg-LeuGluCysSerGlyAlaIleSerAlaHisCysLysLeuArgLeuProGlySerA 458
QY 1062 GYATCTCTGCTAGCCCTCCTAGTACGTGAATTAACAAGCGTGCCACCACCTGCTT 1121
Db 458 rHisSerProAlaSerAlaSerGlnValAlaGlyThrThrGlyAlaArg-ThrThrPro 477
QY 1122 GGCTAATTTTGTATTATTTAGCAGAGATGGGGTTTACCATGTTGCCAGGCTGGTCTC 1181
Db 478 Gly***PhePheValPheLeuValGluThrGlyPheHisArgGlySerGlnaspGlyLeu 497
QY 1182 AAACCTCTGACCTCAAGTGATGCTCCCTGCTCAGTCTCCCAAGTGCTGGAATTATAGGC 1241
Db 498 AspLeuLeuThrSer*****SerAlaArgLeuGlyLeuProLysCysTyrAspTyrArgA 517
QY 1242 GTAGTCACTGTGCTGGCCGATTAAGTCTATTTCTTATTTCTTATTTGCTATATCCCATCT 1301
Db 517 rGlu--ThrAlaProGly*****PheLeuPhePheVal----- 530
QY 1302 AGAGCAGTGTCTGACATATAGTAGTGTCTCAATAAATAATTGATGACACAGCCTAGA 1361
Db 530 ----- 530
QY 1362 TATAAACTTTCTTTTCTTTTAAACAATCTTGACAACTTTGCAGA 1410
Db 531 -----PheIlePhePheArgAspGlyValSerLeuCysArg 542
RESULT 7
E40201
artifact-warning sequence (translated ALU class E) - human
C;Species: Homo sapiens (man)
C;Date: 31-Mar-1992 #sequence_revision 11-Aug-1995 #text_change 19-May-2000
C;Accession: E40201
R;Claverie, J.M.
personal communication, 1992
A;Reference number: A40201
A;Accession: E40201
A;Molecule type: DNA
A;Residues: 1-597 <CLA>
A;Cross-references: UNIPARC:UPI000017CECC
R;Claverie, J.M.
Genomics 12, 838-841, 1992
A;Title: Identifying coding exons by similarity search: Alu-derived and other potential
A;Reference number: A40200; MUID:92241891; PMID:1572661
C;Contents: annotation
C;Comment: This "warning" entry is a conceptual translation in all 6 reading frames of a
in-frame stop codons are shown as 'X'.
C;Comment: Any significant similarity of a predicted protein sequence to a portion of th
Alignment Scores:
Pred. No.: 2.13e-20 Length: 597
Score: 276.50 Matches: 154
Percent Similarity: 37.2% Conservative: 35
Best Local Similarity: 30.3% Mismatches: 136
Query Match: 6.2% Indels: 186
DB: 4 Gaps: 18
US-10-607-806-1-G7328-T9182_COPY_7000_9500 (1-2501) x E40201 (1-597)
QY 1261 GGCCAGGCACAGTGTACTCAGCCCTATATATCCAGCACTTTGGGAGACTGAGCGGAGGAGA 1202
Db 1 GlyGlnValTrp***LeuMetProValIleProAlaLeuTrpGluPheLysAlaGlyArg 20
QY 1201 TC-ACCTGAGTCAAGAGTTTGAGACCGAGCTGGGCAACATGTTAAACCCCATCTCTGC 1143
Db 21 IleThr***GlyTrpGluPheGluThr-SerLeuThrAsnMetGluLysProHisLeuTyr 40
```

```
QY 1142 TAAATAACAAAAATTAGCAAGCATGGTGTGCACGCTGTGTAATTCAGCTACTCAGG 1083
Db 41 ***LysTyr--LysIleSerTrpAlaTrpTrpArgMetProValIleProAlaThrArg 60
QY 1082 AGGCTTAPGCGAG---GAGAATCRCTTGAACCCGGGAGGAGCGAGGTTGCGAGTGTG--- 1029
Db 60 luAla-GluThrGlyGluAsnCysLeuAsnProGlySerLysValCysGlyGluIleVal 79
QY 1028 -----AACGGAGGTGAGGTGATGA-AGTGAGACTCTCTCAAAAAGAAAGGAAGAA 976
Db 80 ProSerTyrSerGlyLeuGlyAsnLysSerLysThrProSerGlnLysLysLysLysLys 99
QY 975 AATAGCAGTAATCCCTGTTCTGTGATCTTACATTTCTATTGGAAGATGGAATGTAAG 916
Db 100 ***AlaArg-----CysGlySer----- 105
QY 915 AATAAGTAGCCGCACATGGTGGCTCACCCCTGTGTAATCCAGACACTTTGGGATGCCAAG 856
Db 106 -----SerCysLeu***SerGlnArgPheGlySerSerArg 117
QY 855 TA-GGAGGATTGCTTGAAGCCAGGAGTTTAAACAAGCCCTGGCCAACTAGTAAAGACCA 797
Db 118 ArgGlyGlySerProGluValGlySerLeuArgProAla***ProThrTrpArgAsnPro 137
QY 796 GTCTCTACAAAAAATTTTAAATAGCAGCGCTGGTGTGCACGCTCGTAGTCCAG 737
Db 138 IleSerThrLysAsnIleLys---LeuAlaGlyArgGlyAlaCysLeu***SerGln 156
QY 736 CTACTTGGGAGGCTGAGGTCAGCAGATCCCTTGGAGCTCAGGAGGTTGAGGCTCAGTGAC 677
Db 157 LeuLeuGlyArgLeuArgGln---GluArgIleAla***ThrGlnGluAlaLysPheAla 175
QY 676 CTATGTTGTCACCACTGTGCTCATGCTGGTGGTGACAGACAA----- 635
Db 634 -----GACCT----- 629
QY 196 LysLys*****ProGlyValValAlaHisAlaCysAsnProSerAlaLeuGlyVal 215
Db 628 -----GTCTTAAACAACAACAACAAAAAACCACAA 599
QY 216 GlnGlyGlyAlaAspHisLeuArgLeuGlyVal***AspGlnLeuAspGlnHisGlyGlu 235
QY 598 ACTGTCAACCTTACAGCTGTGTGAT---CTTGGGCAATAGCTCAGCCCTCTCGAGCCT 542
Db 236 ThrProSerLeuLeuLysIle***Asn***LeuGlyValValAlaHisAlaCysAsnPro 255
QY 541 CAGTTTCTTATCTGTAAAGTGAAGATAATTGTACCTATCCCATAGTGTGTTCTGAGGA 482
Db 256 -SerTyr-----SerGlyGly** 261
QY 481 TTAAGAGACACTGCCCCAGAACCAAGAAAATTAACATACTAATCATGGCTGTGTTACTA 422
Db 261 *AspArgArgGluLeuGluProArgLysGlnSerLeuArg***AspCysAla--Ile 280
QY 421 TTATTTCCCGCCGCTCTACTGAGAACCACTAGAAATTCATAGTCAAG----- 374
Db 281 ValLeuArgProGlyGlnGlnGluGlnAsnSerValSerLysLysLysLysLysLysLys 300
QY 373 -----GAAGGGATAAACCTACTCTACAGGTGATTTGCCAGCCAGCAGACAC 329
Db 301 PhePhePhePhe***AspGlyValLeuLeuLeuLeu-----ProArgPro 315
QY 328 GAGTATGAATAGTGTGGGTGTACGGKKTGTCTCCAGCAGAAATTTACAGCTGTCCAGCTTC 269
Db 316 GlutylAsp-----GlyThrIleSerProGlnThrLeuLeuProGlyPhe 330
QY 268 TTGGCCTGGTCRTAGCAGTTGTTCATGTGTGTCGACGACACCTGGAAAGTGGGAGGACAG 209
Db 330 ----- 330
QY 208 TGAGATAGGAGTAAGTGCAGAGCAATAGGTGAGCCCTCACCTGCCCACTCTCAGGAACAG 149
```

Db 331 -----LysGlnPheSerProValSerAlaSerArgVal 341
QY 148 GTGGGAGTACTTCGCCAAGATGCTTCAGGAGAATGATCTATGGCTTGAAAAATATA 89
Db 342 AlaGlyIleThrGlyMetArg-----HisHis 350
QY 88 AGTCCTTTTATCATATAT-----TTATTGAAA----- 62
Db 351 AlaGlnLeuIleLeuTyrPhe*****ArgTrpGlyPheSerMetLeuValVal 370
QY 62 ----- 62
Db 371 SerAsnSerGlnProGlnValIleArgProAlaLeuAsnSerGlnSerAlaGlyIleThr 390
QY 61 -----GCATTTTTTTTTTTTGGAGACAGAGTTGCTC--- 32
Db 391 GlyMetSerTyrHisThrTrp*****PhePhePheGluThrGluPheCysSer 410
QY 31 TGTATCCAGGC 20
Db 411 CysCysProGly 414

RESULT 8

S41044
Chromosomal protein - human
C;Species: Homo sapiens (man)
C;Date: 19-May-1994 #sequence_revision 10-Nov-1995 #text_change 05-Oct-2004
C;Accession: S41044
R;Yeo, J.P.; Alderuccio, F.; Toh, B.H.
Nature 367, 288-291, 1994
A;Title: A new chromosomal protein essential for mitotic spindle assembly.
A;Reference number: S41044; MUID:94166884; PMID:8121495
A;Accession: S41044
A;Status: preliminary
A;Molecule type: mRNA
A;Residues: 1-418 <YEO>
C;Cross-references: UNIPROT:P49646; UNIPARC:U01000013C29A; GB:I26953; NID:G537529; PIDN:
C;Superfamily: very very hypothetical protein RMSA-1
C;Keywords: chromosomal protein

Alignment Scores:

Pred. No.: 1.14e-19 Length: 418
Score: 269.50 Matches: 122
Percent Similarity: 37.0% Conservative: 51
Best Local Similarity: 26.1% Mismatches: 107
Query Match: 6.1% Indels: 188
DB: 2 Gaps: 20

US-10-607-806-1-G7328-T9182_COPY_7000_9500 (1-2501) x S41044 (1-418)

QY 1428 TTCTGCTTTTCTACTATACCTTGTTATGACTTTTTCATATTGCTTCAACCTTTATG 1487
Db 4 PheCysLeuIleLeuPheAsnLeu-----PheAsnPheCys-----LeuIle 17
QY 1488 TTACTGTGTTTTTCATATG-----TACTATTTTAGT----- 1517
Db 18 LeuSerPheLeuThrCysIleAlaTyrProValLeuTyrPhePheSerPheLeuAsnSer 37
QY 1518 -----CACTGAATAATATG----- 1532
Db 38 GlnThrArgSerLeuLysLeuPheLeuTyrPheLysIleGluIleTrpSerGlnGlyIle 57
QY 1533 CTTAATTTGCTTATACATCTCTCTGCTCCACTTTAGAGCCCAAAATTACAATCTG---- 1589
Db 58 ValArgPheLeuLeuArgPheGlyLeuThrThrPheLysGluArgPheThrSerLeuIle 77
QY 1589 ----- 1589
Db 78 LeuLeuAsnAsnMetHisGlnMetIlePheProMetValIleTyrIleSerLysLeuCys 97
QY 1590 -----ATGAAGCTATGAACCTCTCTCCCAAGAGAAATACACACAC 1631
DB: 3 Gaps: 20

Db 98 IlePheHisPheTrpHisLeuValLeuMetAspLeuValProArgGlnGlnArgSer-I1 117
QY 1632 ACACACTACACACACAGT 1691
Db 117 eileThrTyrSerLeuValPheAlaIleIleSerGlnLysLysLysLysLysLysLys 137
QY 1692 GAGGATGTTTGTTCATATTT----- 1710
Db 137 sLysAsnAsnIleArgIleIleLeuPheLeuProGlnAlaHisGlyArgAspPheTyrVa 157
QY 1711 -----AATTAAAAATAACTAGCTGGGCACAGTACTCAAGCTGTAAAC 1754
Db 157 lProIleLeuProPheThrGlnSerTyrValAspTrpGlyArgTrpLeu----- 173
QY 1755 CACAGTACTTTGGAAGTCCCAAGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGG 1814
Db 174 -----IleTrpGluAlaLysAlaGlyGluSerLeuGluValArgSerSerArgProAl 191
QY 1815 CTGGTCAATATGTTGAAACCCCTACTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 1874
Db 191 aSerGlnSerArgArgAsnSerValSerThrLysAsnIleLysIleSer----- 207
QY 1875 ATGCATGCTCTGA----- 1887
Db 208 -----ProValSerThrLysAsnIleLysIleSerGlnThrTrpTyrLeuPheGlyG 225
QY 1888 -----GTCCAGCTACTCGGAGGCTGAGGCAAGAGAGATTCCTTGAACCTGG 1934
Db 225 yValHisLeuLeuValProThrThrArgAspAlaGluAlaGlyGluLeuHisAspPro-G 245
QY 1935 GAGGACAGAGTTCAGTGCAGCCGAGTCCACCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 1994
Db 245 lGlyArgGlyCysAsnGluLeuArgSerCysHisCysThrProAlaTrpValThrSer- 264
QY 1995 GAGACTCTATCTCAAAAAATAATAATAATAATAATAATAATAATAATAATAATAATAA 2054
Db 265 -----GluThrValSerLysLys-----LysLysLysLysLysLysLysLysLys 282
QY 2055 TAAGATTCTCTGAAGGTAACGAGA-----GATACGTAAATATATATGTAATAAAA- 2101
Db 282 eutThrCysIleAsnAlaSerThrLeuPheHisValLeuThrArgIlePheCysTyrLysG 302
QY 2102 -----GTTTAAATGCATTTTAACTGTAATCTTATTTTGTATTTAT- 2137
Db 302 lnlYsTyrProLeuTrpLeuLeuIleLysArgLeuSerValAsnLeuGlnLysGlnGlnG 322
QY 2138 -----TTTGGTTATAAAAGTAAACAGCCAAAAGT-----AATGCAACTTCAAACT 2183
Db 322 luAspLeuGlnArgLysLysLysGlnArgLysGlnGluHisProGluLeuThrAlaGlnA 342
QY 2184 CTACATAAATATCTATTATGGAAGTGGAGGCAATCTAATCTTACTACTACCCCAAGATAA 2243
Db 342 snThrValLeuLysMetLeuLeuGlnGly---ValLeuIleAsnLeuTyrProThrVal 361
QY 2244 CCAGTTTACATATCTCTCCAGATTTTGGGGCATACACTAGCTTTTTTTTGGGAAAAAT 2303
Db 361 hrLeu----- 362
QY 2304 TTCCATGTGCAGGCATACCTAATTTTCTAAATGTCTATGTAGTATTCATTTAAGGATG 2363
Db 362 ----- 362
QY 2364 TTCCATATTTTAAATAATACATGCTTTAAAGTAGAGAACTAGGTGGCATGGTGGCTC 2423
Db 363 -----AsnArgProGlyValValAlaIle 370
QY 2424 AGCCTGTATCCAGCAGCTTTGGAGGCGCGCAAAATGGATCATCTTGGAGTCCGAGTT 2483
Db 370 isAlaCysIleProAlaLeuTrp---AlaGluGlyGly-ValHisLeuAsnProGly---- 387
QY 2484 TGAGACCCAGCTGGACAA 2501
Db 388 -----GlnProGlyGln 391

Alignment Scores:

Pred. No.: 4,1e-19 Length: 196
Score: 264.00 Matches: 61
Percent Similarity: 68.2% Conservative: 12
Best Local Similarity: 57.0% Mismatches: 23
Query Match: 6.0% Indels: 12
DB: 2 Gaps: 3

US-10-607-806-1-G7328-T9182_COPY_7000_9500 (1-2501) x I38022 (1-196)

QY 1787 CACTTGAGGTGAGAGTTCGAGACCGCCGTCGTAATATGTTGTAACCCCTACTCTCTACTA 1846
Db 3 HisLeuArgSerGlyValGlnAspPyrProGlyGlnHisGlyLysIleProSerLeuLeu 22

QY 1847 AAAATACAAAATTAGCTGGGTGTAGTGATGTCATCCCTGTAGTCCAGTACTCGGGAGG 1906
Db 23 LysIleGlnGluLeuAlaGlyHisGlyArgCysLeuGlnSerGlnLeuLeuArgArg 42

QY 1907 CTGAGGCAAGAAATTCCTTGAACCTGGGAGGCGAGGTTGCAGTGGCCGAGATCCAC 1966
Db 43 LeuArgGlnGluAsnHisLeuAsnSerGlyGlyArgGlyCysSerGluProLysSerHis 62

QY 1967 CACTGCACTCCAGCTCGGCGACACGCGACTCTATCTCAAAAAATAAATAATAA 2026
Db 63 LeuCysIleProAlaIrpValThrGluGlyAspSerValSerLysGln-----AsnLys 80

QY 2027 ATAAGGATCGGAG-----AGAAACAAACTAATAAGATT----- 2061
Db 81 ThrLysAsnGlnGlnHisLeuArgAsnAsnThrLysLysSerAsnSerCysIleIleGly 100

QY 2062 ---CCTGAAGTAAGCAGAGA 2079
Db 101 GlyProGluGlyGluGluLys 107

RESULT 11

PSHU

N;Alternate names: [phospholipase A2 (EC 3.1.1.4) IB precursor [validated] - human
C;Species: Homo sapiens (man)
C;Date: 17-Mar-1987 #sequence_revision 31-Mar-1993 #text_change 16-Aug-2004
C;Accession: C25793; A25793; A00733; A21566
R;Seilhamer, J.J.; Randall, T.L.; Yamanaka, M.; Johnson, L.K.
DNA 5, 519-527, 1986
A;Title: Pancreatic phospholipase A2: isolation of the human gene and cDNAs from porcine
A;Reference number: A25793; MUID:87132925; PMID:3028739
A;Accession: C25793
A;Molecule type: DNA
A;Residues: 1-148 <SEI2>
A;Cross-references: UNIPROT:P04054; UNIPARC:UPI0000048FD9; GB:M22970
A;Accession: A25793
A;Molecule type: mRNA
A;Residues: 1-148 <SEI>
A;Cross-references: UNIPARC:UPI0000048FD9; GB:M21055
A;Experimental source: lung
R;Gratzioli, R.; Dijkman, R.; Dutilh, C.E.; Van der Ouderaa, F.; De Haas, G.H.; Figarella
Eur. J. Biochem. 122, 111-117, 1982
A;Title: Studies on phospholipase A2 and its enzyme from human pancreatic juice. Cata
A;Reference number: A91114; MUID:82138816; PMID:7060561
A;Accession: A00733
A;Molecule type: protein
A;Residues: 16-22 <VER1>
A;Cross-references: UNIPARC:UPI000014C3BF
R;Verheij, H.M.; Westerman, J.; Sternby, B.; De Haas, G.H.
Biochim. Biophys. Acta 747, 93-99, 1983
A;Title: The complete primary structure of phospholipase A2 from human pancreas.
A;Reference number: A90651; MUID:83283533; PMID:6349696
A;Accession: A21566
A;Molecule type: protein
A;Residues: 23-143, 'YS', 146-147 <VER2>
A;Cross-references: UNIPARC:UPI0000172758
C;Genetics:
A;Gene: GDB:PLA2G1B; PLA2A; PLA2; P
A;Cross-references: GDB:120715; OMIM:172410

A;Map position: 12q23-12qter
A;Introns: 12/1; 65/2; 108/1
C;Function:

A;Description: catalyzes hydrolysis of 1,2-diacyl-sn-glycero-3-phosphocholine to 1-acyl
A;Note: the reaction is strongly enhanced when the phospholipid is condensed into a mic
C;Superfamily: Phospholipase A2
C;Keywords: calcium; carboxylic ester hydrolase; lipid degradation; lipid digestion; me
F;1-15/Domain: signal sequence #status predicted <SIG>
F;16-22/Domain: activation peptide #status experimental <ACT>
F;23-148/Product: phospholipase A2 IB #status experimental <MAT>
F;33-99,49-146,51-67,68-127,73-120,83-113,106-118/Disulfide bonds: #status predicted
F;50,52,54,71/Binding site: calcium (Tyr, Gly, Asp) #status predicted
F;70,121/Active site: His, Asp #status predicted

Alignment Scores:

Pred. No.: 1.73e-18 Length: 148
Score: 258.00 Matches: 55
Percent Similarity: 59.8% Conservative: 6
Best Local Similarity: 53.9% Mismatches: 12
Query Match: 5.8% Indels: 29
DB: 1 Gaps: 3

US-10-607-806-1-G7328-T9182_COPY_7000_9500 (1-2501) x PSHU (1-148)

QY 113 ATTTCTCTCAAGCATCTTTGGCGA-----AGT 139
Db 19 IleSerProArgAlaValTrpGlnPheArgLysMetIleLysCysValIleProGlySer 38

QY 140 CATCCACACTG-----TTCTGAGAGTGGCGAGGTGAGGGCTG 178
Db 39 AspProPheLeuGluTyrAsnAsnTyrGlyCysTyrCysGlyLeuGly-GlySerGlyTh 58

QY 179 ACCTATTGCTCTGCACCTTACTCTCATCTCAGCTGTCCCTCCCACTTTCCAGGTTGCCA 238
Db 58 rProValAsp-----GluLeuAspLysCysCysGly 68

QY 239 GACACATGACAACTGCTAYGACCGCCAGCAAGCTGGACAGCTGTAAATTTCTGCTGGA 298
Db 68 nThrHisAspAsnCysTyrAspGlnAlaLysLysLeuAspSerCysLysPheLeuLeuAs 88

QY 299 CMMCCGTACACCCACACTTATTCATCTGCTGCTGGCTGGCAATCACTGTAGCAG 358
Db 88 pAsnProTyrThrHisThrTyrSerTyrSerCysSerGlySerAlaIleThrCysSerSe 108

QY 359 TAGG 362

Db 108 rLys 109

RESULT 12

B40201

artifact-warning sequence (translated ALU class B) - human

C;Species: Homo sapiens (man)

C;Date: 31-Mar-1992 #sequence_revision 11-Aug-1995 #text_change 19-May-2000

C;Accession: B40201

R;Claverie, J.M.

personal communication, 1992

A;Reference number: A40201

A;Accession: B40201

A;Molecule type: DNA

A;Residues: 1-301 <CLA>

A;Cross-references: UNIPARC:UPI000017CEC9

R;Claverie, J.M.

Genomics 12, 838-841, 1992

A;Title: Identifying coding exons by similarity search: Alu-derived and other potential

A;Reference number: A40200; MUID:92241891; PMID:1572661

A;Contents: annotation

C;Comment: This "warning" entry is a conceptual translation in all 6 reading frames of
in-frame stop codons are shown as 'X'.

C;Comment: Any significant similarity of a predicted protein sequence to a portion of t

Alignment Scores:

Pred. No.: 1.88e-17 Length: 301

Score: 248.50 Matches: 80
Percent Similarity: 43.7% Conservativity: 17
Best Local Similarity: 36.0% Mismatches: 74
Query Match: 5.6% Indels: 52
DB: 4 Gaps: 8

US-10-607-806-1-G7328-T9182_COPY_7000_9500 (1-2501) x B40201 (1-301)

Qy 1261 GGCCAGGACAGTGCATCAGCCCTATATTCAGCAGCTTTGGGAGACTGAGGAGGCGAGA 1202
Db 1 GlyArgAlaTrpTrpLeuThrSerValIleProAlaLeuTrpAspAlaGluValGlyGly 20
Qy 1201 TCACCTGAGTTCAGGAGCTTGGAGCAGCTGGGCAACATGGTAAACCCCATCTGCT 1142
Db 21 LeuLeuGluThrArgSerSerArgProAlaTrpAlaThr***GlnAspProSerSerile 40
Qy 1141 AAAAATACAAAAATTA-----GCCAAGCATGTGTGTCAGCGCTTTGTA 1098
Db 41 LysIleLys***Asn***LeuGly*****AlaGlyHisGlyGly***HisLeu*** 60
Qy 1097 TTCAGCTACTCAGGAGCTTAGCAGGAGAAATCCTTGAACCCGGGAGGAGGTTGC 1038
Db 61 SerGlnHisPheGlyMetLeuLys***GluAspCysLeuArgProGly----- 76
Qy 1037 AGTCAGCTGAACGAGGTTGAGGTGATGACGTGAGACTCTCTCAAAAAGAAAGGAG 978
Db 77 -----ValGlnAspGlnLeuGlyGlnHisSerLysThrProAlaLeu***Lys***Asn 94
Qy 977 AAAATAGCGAGTAATCCCTGTCTGTGGATCTTACATCTTATTTGGAAGATGGAATGAA 918
Db 95 LysIleSer-----TnpVal*****----- 102
Qy 917 AGAAATAAGTAGCGCACATGGTGGCTCACCCCTGTATATCCAGCAGCTTTGGGATGCCAA 858
Db 103 -----GlnGlyMetValAlaAsnIleCysAsnProSerThrLeuGlyCys*** 118
Qy 857 GGTAGGAGGATGCTTGAAGCAGGAGTTTAAAGCAAGCCTGGCCCAACGTAGTAAGACCC 798
Db 119 SerArgArgIleAla***AspGlnGluPheLysThrSerLeuGlyAsnIleAlaArgPro 138
Qy 797 AGTCTCTACAAA-----AAAAAT-----TTTTTA 774
Db 139 GlnLeuThrLysAsnLysIleLysLeuAlaGly*****HisProAlaAsnPheIle 158
Qy 773 ATTAGCAGGCGTGGTGTGCACGCTGTAGTCCAGCT-----ACTGGGAG 726
Db 159 LeuPheLeu***SerTrpGlyLeuAlaMetLeuProLysLeuValLeuAsnSerTrpSer 178
Qy 725 GCTGAGGTCAGCAGATCCCTTGTAGCTCAGGAGGTTGAGGCTGACGTGACCTATGTTGCA 666
Db 179 GlnAlaIle-----LeuLeuLeuGlnHisProLysValLeu 190
Qy 665 CCACTG 660
Db 191 GlyLeu 192

RESULT 13
B40201
artifact-warning sequence (translated ALU class B) - human
C;Species: Homo sapiens (man)
C;Date: 31-Mar-1992 #sequence_revision 11-Aug-1995 #text_change 19-May-2000
C;Accession: B40201
personal communication, 1992
R;Claverie, J.M.
A;Reference number: A40201
A;Accession: B40201
A;Molecule type: DNA
A;Residues: 1-301 <CLA>
A;Cross-references: UNIPARC:UPI0000017CEC9
Genomics 12, 838-841, 1992
A;Title: Identifying coding exons by similarity search: Alu-derived and other potential
A;Reference number: A40200; MUID:92241891; PMID:1572661

A;Contents: annotation
C;Comment: This "warning" entry is a conceptual translation in all 6 reading frames of or
in-frame stop codons are shown as 'X'.
C;Comment: Any significant similarity of a predicted protein sequence to a portion of thi

Alignment Scores: 1.69e-16 Length: 301
Pred. No.: 239.50 Matches: 81
Score: 239.50
Percent Similarity: 48.6% Conservativity: 23
Best Local Similarity: 37.9% Mismatches: 67
Query Match: 5.4% Indels: 44
DB: 4 Gaps: 10

US-10-607-806-1-G7328-T9182_COPY_7000_9500 (1-2501) x B40201 (1-301)

Qy 667 GCACCATAGTGCATCAGCCCTCAACCTCCTGAGCTCAAGGATCTGTGACCTCAGCC 726
Db 113 SerThrLeuGlyCys***SerArgArgIleAla***AspGlnGluPheLysThrSerLeu 132
Qy 727 TCCCAAGTAGCTGG-----ACTACGAGCGTGAC 756
Db 133 GlyAsnIleAlaArgProGlnLeuThrLysAsnLysIleLysLeuAlaGly***** 152
Qy 757 CACCACGCTGGCTAATTAATAAATTTTCTGAGAGACTGGGTCTTACTAGCTTGGCC 816
Db 153 HisProAlaAsnPheIle-----LeuPheLeu***SerTrpGlyLeuAlaMetLeuPro 170
Qy 817 AGCTTGTCTTAAACTCCTGGCTTCAAGCAATCCTCTACCTTGGCATCCCAAGTCTG 876
Db 171 LysLeuValLeuAsnSerTrpSerGlnAlaIleLeuLeuGlnHisProLysValLeu 190
Qy 877 GATTAACAGGGGTGAGCCACCATG---TGCGGCTACTATTCTTTTACATCTTCATCTTC 933
Db 191 GlyLeuGlnMetLeuAlaThrMetProCys----- 200
Qy 934 CAATAGAATTAAGATCCACAGAACAGGAGTACTGCTATTCTTCTTCTTTCTTTTGT 993
Db 201 -----*****ThrGln-----LeuIleLeuPheTyr-----PheTyr 212
Qy 994 AGACAGAGTCTCATCTTCATCACCTCAACCTCGTTCAGCTCACTGCAACCTCTGCTCC 1053
Db 213 ArgAlaGlyValLeuLeuCys---CysProSerTrpSer***-----ThrPro 227
Qy 1054 GGTTTCAGGATCTCTCCCTAGCCCTCCTGAGTCTGGAATTAACAGCGTGACCA 1113
Db 228 GlyLeuLysGlnSerTyrPheSerIleProLysCysTrpAspTyrArgCys***Pro 247
Qy 1114 CCATGCTTGGCTAA-----TTTTTCTATTATTTAGCAGATGGG 1154
Db 248 ProCysProAla*****ProSer***PheTyrPheIlePheIle---GluLeuGly 266
Qy 1155 TTTTACCATGTTCCAGGCTGTCTCAAACTCCTGACCTCAAGTGAATCTGCTGCTCA 1214
Db 267 SerCysTyrValAlaGlnAlaGlyLeuGluLeuValSerSerAsnProProThrSer 286
Qy 1215 GTCTCCCAAGTCTGGAATTAATATAGCGTCACTGCTGCC 1256
Db 287 AlaSerGlnSerAlaGlyIleThrAspValSerHisAla 300

RESULT 14
D40201
artifact-warning sequence (translated ALU class D) - human
C;Species: Homo sapiens (man)
C;Date: 31-Mar-1992 #sequence_revision 11-Aug-1995 #text_change 19-May-2000
C;Accession: D40201
R;Claverie, J.M.
personal communication, 1992
A;Reference number: A40201
A;Accession: D40201
A;Molecule type: DNA
A;Residues: 1-579 <CLA>
A;Cross-references: UNIPARC:UPI0000017CECB

R:Claverie, J.M.
 Genomics 12, 838-841, 1992
 A;Title: Identifying coding exons by similarity search: Alu-derived and other potential
 A;Reference number: A40200; MUID:92241891; PMID:1572661
 A;Contents: annotation
 C;Comment: This "warning" entry is a conceptual translation in all 6 reading frames of c
 in-frame stop codons are shown as 'X'.
 C;Comment: Any significant similarity of a predicted protein sequence to a portion of th

Alignment Scores:
 Pred. No.: 2,03e-16 Length: 579
 Score: 239.00 Matches: 125
 Percent Similarity: 34.7% Conservative: 24
 Best Local Similarity: 29.1% Mismatches: 121
 Query Match: 5.4% Indels: 160
 DB: 4 Gaps: 14

US-10-607-806-1-G7328-T9182_COPY_7000_9500 (1-2501) x D40201 (1-579)

QY 11 TGCTGTCCAGCTGGTAAACAGAGCAACTCTGTCTCAAAAAAATAATGCTTCAATAA 70
 |||||
 Db 299 CysCysProGlyTrpSerThrGlu----- 306
 QY 71 ATATATGATAAAGGACTTATATATTTTCAAGCCATAGGATCAATTTCTCTGGAAGCATCT 130
 |||||
 Db 307 -----ValGlnSer 309
 QY 131 TGGCGAAGTCATCCCACTCTGTTCTGAGAGTGGGAGGTGAGGGCTGACCTATTGCTCT 190
 |||||
 Db 310 TrpLeuThrGluPro***Leu----- 316
 QY 191 GCATTACTCTATCTCAGCTGTCCT-----CCCACTTTCCAGGTGTCGCCAGACACA 244
 |||||
 Db 317 -----ProArgSerSerAppProAlaProAlaPhe***Val-AlaGlyThrTh 333
 QY 245 TGAACAAGTCTAYGACCAAGCAAGAGTGGAGCTGTAAATTTCTGCTGGACAMWCC 304
 |||||
 Db 333 rGlyLeuCysTyGln----- 339
 QY 305 GTACACCCACCTATTATCATCTCTGCTCTGCTGGCAATCACCTGTAGCAGTAGGTT 364
 |||||
 Db 339 eTrpLeuPheValSerPheCysArgGlyGlyValSerThrCysCys----- 356
 QY 365 TATCCCTCTCTGACCTATGAATCTAGTTGTTCTCAGTAGGCGGGGGAATAATAG 424
 |||||
 Db 357 -----ProGlyCysSer***Th 362
 QY 425 TAACAAC---AGCCATGATTAGTGTAAATTTCTTGGTCTGGGCAAGTCTCTCTTTAA 481
 |||||
 Db 362 rProGluLeuGluHisThrProAlaMetAlaProLysValLeuGlyLeuGlnAla----- 380
 QY 482 TCCTCAGAACACACTATGGGATAGGTACAAATTTCTCTCACTTAACAGATAAGAAACTG 541
 |||||
 Db 380 ----- 380
 QY 542 AGGCTCAGAGGCTGAGCTATTGCCCCAAGATCACACAGCTTGTAAAGTGTGACAGTTG 601
 |||||
 Db 381 -----Cys-----SerHisAla-----Tr 385
 QY 602 GGTATTTTGTGTTGTAGACAGAGGTCTGCTCTGTCAACCCAGGCTGAGCACA 661
 |||||
 Db 385 P*****-----PheGlnAlaGlySerTrpTyrrValAlaGlnAlaGlyValGI 402
 QY 662 GTGGTGAACCAATAGTCTAGTCAAGCTCAACTCTGAGCTCAAGGATCTCTGACCT 721
 |||||
 Db 402 nArgCysAsnHisGlySerLeu-SerLeuAspPheProGlyGlnValLeuLeu-Leu-Pro 421
 QY 722 CAGCTCCCAAGTAGTGGGACTACGAGGTGACCCAGGCTGGC---TAATATAAA 778
 |||||
 Db 422 GlnProPheGlu***LeuGlyLeuGlnAlaTyAlaThrArgSerGly***PheLeuTy 441
 QY 779 AAATTTTGTAGAGACTGGGTCTTACTACGTTGGCCAGGCTGTCTTAAACTCCTGGC 838

Db 442 Leu-PhePheValGluValGlyPheArgHisValAlaGlnAlaValLeuGluLeuLeuSe 461
 |||||
 QY 839 TTCAAGCAATCCCTTACCTTGGCATCCCAAGTGTGGATACAGGGGTGACCCACCA 898
 |||||
 Db 461 rTrpSer-IleHisLeuProTrp-LeuLeuLysCysTrpAspTyArgArg--AlaAlaM 480
 |||||
 QY 899 TGTGCGGCTACTTATTCTTTTACATTCCATCTTCCAATAGAAATGTAAAGATCCACAGAC 958
 |||||
 Db 480 etLeuGly----- 482
 QY 959 AGGATTACTGCCTATTCTTCTTTTGTAGACAGAGTCTCATCTTCATCACCCTC 1018
 |||||
 Db 483 -----PheArgGlnGlyLeuGlyMetLeuProA 495
 QY 1019 AACCTCGTTACGCTCACTG-----CAACCTCTGCCTCCCGGGTTCAGAGTAT 1066
 |||||
 Db 495 rLeuGluTyArgGlyAlaIleMetAlaHis***AlaLeuThrSerGlnValLys***S 515
 QY 1067 TCTCCTCGCTAAGCCTCTGAGTAGCTGGAATACAAAGCGTGACACCATGCTTGGCTA 1126
 |||||
 Db 515 erSerCysProSerLeuLeuSerSerTrpAspTyArgLeuMetLeuProAspLeuAla 535
 QY 1127 ATTTT---TTGATATTTTAGCAGATGGGGTTTTTACCATGTTGCCAGGCTGTCTCAA 1183
 |||||
 Db 535 snPheCysIlePhePheLeu***ArgTrpGlyPheAspMetLeuProArgLeuPheLeu 555
 QY 1184 ACTCCTGACCTCAAGTATCTGCTGCTCAGTCTCCAAAGTGTGGAAATTTATAGCGGT 1243
 |||||
 Db 555 snSer***AlaGlyAlaTyThrCysHisGlySer***SerAlaGlyIleThrGlyVa 574
 QY 1244 GAGTCACTGT 1253
 |||||
 Db 574 GlnProCys 577

RESULT 15

A46010
 X-linked retinopathy protein (C-terminal, clone XEH.8c) - human (fragment)
 C;Species: Homo sapiens (man)
 C;Date: 21-Sep-1993 #sequence_revision 18-Nov-1994 #text_change 09-Jul-2004
 C;Accession: A46010
 R;Wong, P.; MacDonald, I.M.; Sood, R.; Smith, C.; Pilon, R.; Tenniswood, M.
 A;Title: Identification and partial characterization of a candidate gene for X-linked r
 Genomics 15, 467-471, 1993
 A;Reference number: A46010; MUID:93224131; PMID:8468040
 A;Accession: A46010
 A;Status: preliminary
 A;Molecule type: nucleic acid
 A;Residues: 1-100 <WON>
 A;Cross-references: UNIPROT:Q07826; UNIPARC:UPI000006E23F; GB:S58722; NID:g299470; PIDN
 A;Note: sequence extracted from NCBI backbone (NCBIN:129339, NCBIP:129340)

Alignment Scores:
 Pred. No.: 8.51e-16 Length: 100
 Score: 232.50 Matches: 51
 Percent Similarity: 75.3% Conservative: 10
 Best Local Similarity: 63.0% Mismatches: 18
 Query Match: 5.3% Indels: 2
 DB: 2 Gaps: 1

US-10-607-806-1-G7328-T9182_COPY_7000_9500 (1-2501) x A46010 (1-100)

QY 2022 TTTATTTATTTTGTAGATAGAGTCTGCTGTGCTGCCAGGCTGGAGTGCAGTGGTGG 1963
 |||||
 Db 2 PhePhePhePhePheGluThrGluSerCysSerValAlaGluAlaGlyValGlnTrpCys 21
 |||||
 QY 1962 GATCTCGGTCACTGCAACTCTGCTCTCCAGTTCACCAAGCAATTCCTTTCCTCCACGCTC 1903
 |||||
 Db 22 AspLeuGlySerLeuLysSerProPro-FroglySerSerAspSerProAlaSerAla 41
 QY 1902 CCGAGTAGCTGGAGCTACAGGATGATCATCACACCCAGCTAAATTTTGTATTTTAGT 1843
 |||||
 Db 41 rArgValAlaGlyIleThrGlyMetHisHisThrGlnLeuIlePheValPheLeuVa 61

GenCore version 5.1.1.7
Copyright (c) 1993 - 2006 Bioceleration Ltd.

OM nucleic - protein search, using frame_plus_n2p model

Run on: February 8, 2006, 10:23:31 ; Search time 115.328 Seconds
(without alignments)
3060.024 Million cell updates/sec

Title: US-10-607-806-1-G7328-T9182_COPY_7000_9500

Perfect score: 4426

Sequence: 1 gtgtgtcactgtgtccag.....tttgagaccagcctggacaa 2501

Scoring table:

BLOSUM62
Xgapop 10.0 , Xgapext 0.5
Ygapop 10.0 , Ygapext 0.5
Fgapop 6.0 , Fgapext 7.0
Delop 6.0 , Delext 7.0

Searched: 2166443 seqs, 705528306 residues

Total number of hits satisfying chosen parameters: 4332886

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Command line parameters:

-MODEL=frame+ n2p.model -DEV=xlp
-Q=/abs/ABSSWEB_spool/WARTIN10607806/runat_08022006_092018_27658/app_query.fasta_1
-DB=UniProt -QFMT=fastan -SUFFIX=n2p.rup -MINMATCH=0.1 -LOOPCL=0 -LOOPEXT=0
-UNITS=bits -START=1 -END=1 -MATRIX=blosum62 -TRANS=human40.cdi -LIST=45
-DOCALIGN=200 -THR_SCORE=pct -THR_MAX=100 -THR_MIN=0 -ALIGN=15 -MODE=LOCAL
-OUTFMT=pco -NORM=ext -HEADSIZE=500 -MINLEN=0 -MAXLEN=2000000000 -HOST=abes02p
-USER=WARTIN10607806 @CGN 1.1.694 @runat_08022006_092018_27658 -NCPU=6 -ICPU=3
-NO MAP -NEG_SCORES=0 -WAIT_DSPELOCK=100 -LONGLOG -DEV_TIMEOUT=120
-WARN_TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5 -FGAPOP=6 -FGAPEXT=7
-YGAPOP=10 -YGAPEXT=0.5 -DSELOP=6 -DELEXT=7

Database : UniProt 05.80.*

1: uniprot_sprot.*

2: uniprot_trembl.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	502.5	11.4	591	1 ALU1_HUMAN	P39188 homo sapien
2	498	11.3	591	1 ALU1_HUMAN	P39188 homo sapien
3	479	10.8	593	1 ALU7_HUMAN	P39194 homo sapien
4	463	10.5	593	1 ALU7_HUMAN	P39194 homo sapien
5	443	10.0	593	1 ALU6_HUMAN	P39193 homo sapien
6	442	10.0	591	1 ALU8_HUMAN	P39195 homo sapien
7	439.5	9.9	591	1 ALU8_HUMAN	P39195 homo sapien
8	425.5	9.6	593	1 ALU6_HUMAN	P39193 homo sapien
9	406.5	9.2	587	1 ALU2_HUMAN	P39189 homo sapien
10	397.5	9.0	375	2 O60448_HUMAN	O60448 homo sapien
11	396.5	9.0	239	2 Q6ZWA9_HUMAN	Q6ZWA9 homo sapien
12	386	8.7	179	2 Q6ZS15_HUMAN	Q6ZS15 homo sapien
13	384.5	8.7	158	2 Q8NAL9_HUMAN	Q8NAL9 homo sapien
14	382.5	8.6	585	1 ALU5_HUMAN	P39192 homo sapien
15	382.5	8.6	587	1 ALU3_HUMAN	P39190 homo sapien
16	381	8.6	241	2 Q6ZTY3_HUMAN	Q6ZTY3 homo sapien

C	17	377.5	8.5	587	1	ALU2_HUMAN	P39189 homo sapien
	18	372.5	8.4	585	1	ALU5_HUMAN	P39192 homo sapien
	19	372	8.4	200	2	Q6ZUK3_HUMAN	Q6ZUK3 homo sapien
C	20	366.5	8.3	156	2	Q8NBH4_HUMAN	Q8NBH4 homo sapien
	21	365	8.2	603	1	ALU4_HUMAN	P39191 homo sapien
	22	363	8.2	314	2	Q6ZTL7_HUMAN	Q6ZTL7 homo sapien
C	23	363	8.2	603	1	ALU4_HUMAN	P39191 homo sapien
	24	361	8.2	146	2	Q6ZRX6_HUMAN	Q6ZRX6 homo sapien
C	25	360.5	8.1	128	2	Q6ZPB2_HUMAN	Q6ZPB2 homo sapien
	26	350.5	7.9	129	2	Q6ZUW8_HUMAN	Q6ZUW8 homo sapien
C	27	344.5	7.8	150	2	Q6ZPA0_HUMAN	Q6ZPA0 homo sapien
	28	344	7.8	123	2	Q6ZMQ3_HUMAN	Q6ZMQ3 homo sapien
C	29	342	7.7	129	2	Q8N9K0_HUMAN	Q8N9K0 homo sapien
	30	339.5	7.7	176	2	Q6ZRS2_HUMAN	Q6ZRS2 homo sapien
C	31	336	7.6	124	2	Q6ZSN0_HUMAN	Q6ZSN0 homo sapien
	32	335.5	7.6	170	2	Q8NAI3_HUMAN	Q8NAI3 homo sapien
C	33	335.5	7.6	587	1	ALU3_HUMAN	P39190 homo sapien
	34	334	7.5	164	2	Q6ZUK0_HUMAN	Q6ZUK0 homo sapien
C	35	334	7.5	375	2	O60448_HUMAN	O60448 homo sapien
	36	333	7.5	137	2	Q6ZS97_HUMAN	Q6ZS97 homo sapien
C	37	331.5	7.5	127	2	Q6ZVM2_HUMAN	Q6ZVM2 homo sapien
	38	329.5	7.4	157	2	Q8N845_HUMAN	Q8N845 homo sapien
C	39	329	7.4	167	2	Q6ZSN3_HUMAN	Q6ZSN3 homo sapien
	40	328.5	7.4	120	2	Q6ZTA3_HUMAN	Q6ZTA3 homo sapien
C	41	328	7.4	140	2	Q6ZP99_HUMAN	Q6ZP99 homo sapien
	42	327.5	7.4	115	2	O9N083_MACFA	O9N083 macaca fasc
C	43	327	7.4	131	2	Q6ZTD3_HUMAN	Q6ZTD3 homo sapien
	44	326	7.4	164	2	Q6ZTS0_HUMAN	Q6ZTS0 homo sapien
	45	326	7.4	334	2	Q6ZRP97_HUMAN	Q6ZRP97 homo sapien

ALIGNMENTS

RESULT 1

ALU1_HUMAN
ID ALU1_HUMAN STANDARD; PRT; 591 AA.
AC P39188;
DT 01-FEB-1995 (Rel. 31, Created)
DT 01-FEB-1995 (Rel. 31, Last sequence update)
DT 10-MAY-2005 (Rel. 47, Last annotation update)
DE Alu subfamily J sequence contamination warning entry.
OS Homo sapiens (human)
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini; Hominidae;
OC Homo.
OX NCBI_TaxID=9606;
RN [1]
RP NUCLEOTIDE SEQUENCE.
RX MEDLINE=95021758; PubMed=7935834; DOI=10.1038/371752a0;
RA Claverie J.-M., Makalowski W.;
RT "Alu alert."
RL Nature 371:752-752 (1994).
RN [2]
RP CONCEPT.
RX MEDLINE=92241891; PubMed=1572661;
RA Claverie J.-M.;
RT "Identifying coding exons by similarity search: alu-derived and other potentially misleading protein sequences.";
RL Genomics 12:838-841 (1992).
[3]
RX ALU FAMILIES CLASSIFICATION.
RX MEDLINE=88333009; PubMed=3138422;
RA Quentin Y;
RT "The Alu family developed through successive waves of fixation closely connected with primate lineage history.";
RL J. Mol. Evol. 27:194-202 (1988).
[4]
RX ALU FAMILIES CLASSIFICATION.
RX MEDLINE=91178815; PubMed=1706781;
RA Jurka J., Milosavljevic A.;
RT "Reconstruction and analysis of human Alu genes.";
RL J. Mol. Evol. 32:105-121 (1991).
CC -!- MISCELLANEOUS: Various analyses indicate that Alu repeats fall

into 8 subfamilies (Ref.3 and Ref.4). Therefore, 8 Alu warning consensus sequences have been constituted that contain all six frames conceptual translations of each of these classes of Alu repeats.

-1- MISCELLANEOUS: Isolated 'X' indicates the presence of a stop codon, 'XXX' is used to separate the various translation phases.

-1- CAUTION: This Alu entry is provided in order to avoid the further pollution of protein sequence databases with Alu-derived amino acid sequences.

-1- CAUTION: Alu repetitive sequences are interspersed in human and primate genomes with an average spacing of 4 kb. Some of them are actively transcribed by pol III. Normal transcripts may contain Alu-derived sequences in 5' or 3' untranslated regions. However, cDNA libraries also contain partial and/or rearranged cDNAs ligated with Alu-derived sequence in any orientation. Although Alu elements (especially situated on the complementary strand) have a great potential to create additional/alternative exons, consideration should be given to the possibility that the presence of an Alu in an open reading frame may have resulted from a cloning artifact or may be due to misinterpretation of sequencing data. This point has been overlooked on several occasions, with the consequence of erroneous Alu-derived amino acid sequences being reported.

-1- CAUTION: Any significant similarity of a putative protein sequence with an Alu-translated entry must be taken as a warning that a part of Alu repeat may have been artifactually included in the coding nucleotide sequence.

This Swiss-Prot entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use as long as its content is in no way modified and this statement is not removed.

EMBL; U14567; -; NOT_ANNOTATED_CDS; Genomic_DNA.

DR KW HYPOTHETICAL PROTEIN.

FT REGION 1 96 Frame-1.

FT REGION 100 195 Frame-2.

FT REGION 199 294 Frame-3.

FT REGION 298 393 Frame-4.

FT REGION 397 492 Frame-5.

FT REGION 496 591 Frame-6.

SQ SEQUENCE 591 AA; 63827 MW; 665D395735519D95 CRC64;

Alignment Scores:

Pred. No.: 3.51e-42 Length: 591

Score: 502.50 Matches: 129

Percent Similarity: 48.6% Conservative: 15

Best Local Similarity: 43.6% Mismatches: 65

Query Match: 11.4% Indels: 87

DB: 1 Gaps: 4

US-10-607-806-1-G7328-T9182_COPY_7000_9500 (1-2501) x ALU1_HUMAN (1-591)

QY 607 TTTTGTGTTGTTAGACAGGGCTTGCTGTGTCACCCAGGCATGACAGTGGT 666

DB 298 PhePheLeu-----AAGGinglyLeuAlaLeuSerProArgLeuGluCysSerGly 314

QY 667 GCACCATAGTGCTACTGCAGCCTCAACCTCTCTGAGCTCAAGGGATCTGCTGACCTCAGCC 726

DB 315 AlaIleThrAlaHisCysSerLeuAspLeuProGlySerSerAspProProAlaSerAla 334

QY 727 TCCCAAGTAGTGGGACTACGAGCGTGCACACACCCCTGGCTAATAAATAATTTT 786

DB 335 SerArgValAlaGlyThrThrGlyAlaArgHisAlaArgLeuIlePheVal--PheP 354

QY 787 TTGTAGAGACTGGGCTCTACTACCTGGCCAGGCTTGCTTTAACTCTGCTCTCAAGCA 846

DB 354 heValGluThrGlyPheHisTyrValAlaGlnAlaGlyLeuGluLeuLeuGlySerSera 374

QY 847 ATCTCTTACCTGGCATCCCAAGTCTGGATTACAGGGGTGAGCCCATGTGGCGG- 905

DB 374 spProProAlaSerAlaSerGlnSerAlaGlyIleThrGlyValSerHisArgAlaArg* 394

QY 905 ----- 905

DB 394 *****PhePhe***AspArgValSerLeuCysArgProGlyTrpSerAlaValAlaA 414

QY 906 -----CTACTTATTTCTTTAC 921

DB 414 rgSerArgLeuThrAlaAlaSerThrSerArgAlaGlnAlaIleLeuLeuProGlnProp 434

QY 922 ATTCCATCTTTCAATAGATCCAGACAGAGGATTAATCTGCTATTTCTTCTT- 980

DB 434 roGlu***LeuGlyLeuGlnAlaArgAlaThrThrProGly***PheLeuTyrPheLeu* 454

QY 980 ----- 980

DB 454 **ArgArgGlyPheThrMetLeuProArgLeuValSerAsnSerTrpAlaGlnValIleL 474

QY 981 -----C 981

DB 474 euProProArgProProLysValLeuGlyLeuGlnAla***AlaThrAlaProGly*** 494

QY 982 CTTTCTTTTGTGACAGAGTCTCATCTCATCACCTCAACCTCCGTTTCAGC----- 1032

DB 494 ****PhePheGluThrGlySerArgSerValAlaGlnAlaGlyVal-GlnTrpArgAsp 513

QY 1033 -----TCACTGCACCTCTGCTCCGGGTTTCAAGYGATTCTCTGCTTAAGCCCTCTG 1086

DB 514 HisGlySerLeuGlnProArgProProGlyLeuLysArgSerSerCysLeuSerLeuPro 533

QY 1087 AGTAGCTGGAATTAACAGCGTCACCCACCATGCTGGCTAATTTTGTATTTTATGACGAG 1146

DB 534 SerSerTrpAspTyrArgArgAlaProProArgProAlaAsn-PheCysIlePheCysAr 553

QY 1147 AGATGGGTTTACCATGTTGCCAGGCTGGTCTCAAACTCTGACCTCAAGTATGCTGC 1206

DB 553 gAspGlyValSerLeuCysCysProGlyTrpSerArgThrProGlyLeuLys***SerSe 573

QY 1207 CTGCTCAGCTCTCCCAAGTGTGGAAATATATAGCGTGAG 1246

DB 573 rArgLeuGlyLeuProLysCysTrpAspTyrArgArgGlu 586

RESULT 2

ID ALU1_HUMAN STANDARD; PRT; 591 AA.

AC P39188;

DT 01-FEB-1995 (Rel. 31, Created)

DT 01-FEB-1995 (Rel. 31, Last sequence update)

DT 10-MAY-2005 (Rel. 47, Last annotation update)

DE Alu subfamily J sequence contamination warning entry.

OS Homo sapiens (Human).

OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

OC Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini; Hominidae;

OC Homo.

OX NCBI_TaxID=9606;

RP [1]

RP NUCLEOTIDE SEQUENCE.

RX MEDLINE=95021758; PubMed=7935834; DOI=10.1038/371752a0;

RA Claverie J.-M., Makalowski W.;

RT "Alu alert.";

RL Nature 371:752-752 (1994).

RL [2]

RP CONCEPT.

RX MEDLINE=92241891; PubMed=1572661;

RA Claverie J.-M.;

RT "Identifying coding exons by similarity search: alu-derived and other potentially misleading protein sequences.";

RL Genomics 12:838-841 (1992).

RL [3]

RP ALU FAMILIES CLASSIFICATION.

RX MEDLINE=88333009; PubMed=3138422;

RA Quentin Y.;

RT "The Alu family developed through successive waves of fixation closely connected with primate lineage history.";

RT

CC acid sequences.

CC -!- CAUTION: Alu repetitive sequences are interspersed in human and

CC primate genomes with an average spacing of 4 kb. Some of them are

CC actively transcribed by pol III. Normal transcripts may contain

CC Alu-derived sequences in 5' or 3' untranslated regions. However,

CC cDNA libraries also contain partial and/or rearranged cDNAs

CC ligated with Alu-derived sequence in any orientation. Although Alu

CC elements (especially situated on the complementary strand) have a

CC great potential to create additional/alternative exons,

CC consideration should be given to the possibility that the presence

CC of an Alu in an open reading frame may have resulted from a

CC cloning artifact or may be due to misinterpretation of sequencing

CC data. This point has been overlooked on several occasions, with

CC the consequence of erroneous Alu-derived amino acid sequences

CC being reported.

CC -!- CAUTION: Any significant similarity of a putative protein sequence

CC with an Alu-translated entry must be taken as a warning that a

CC part of Alu repeat may have been artifactually included in the

CC coding nucleotide sequence.

CC -----

CC This Swiss-Prot entry is copyright. It is produced through a collaboration

CC between the Swiss Institute of Bioinformatics and the EMBL outstation -

CC the European Bioinformatics Institute. There are no restrictions on its

CC use as long as its content is in no way modified and this statement is not

CC removed.

CC -----

CC EMBL; U14572; -- NOT ANNOTATED_CDS; Genomic_DNA.

CC InterPro; IPR002218; GIDA.

CC ProDom; PD003738; GIDA; 1.

CC Hypothetical protein.

CC REGION 1 97 Frame-1.

CC REGION 101 196 Frame-2.

CC REGION 200 295 Frame-3.

CC REGION 299 395 Frame-4.

CC REGION 399 494 Frame-5.

CC REGION 498 593 Frame-6.

CC SEQUENCE 593 AA; 64640 MW; 136EF344AACD12A2 CRC64;

CC -----

CC Alignment Scores:

CC Pred. NO.: 4.91e-36 Length: 593

CC Score: 443.00 Matches: 113

CC Percent Similarity: 56.6% Conservative: 7

CC Best Local Similarity: 53.3% Mismatches: 61

CC Query Match: 10.0% Indels: 31

CC DB: 1 Gaps: 2

CC -----

CC US-10-607-806-1-G7328-T9182_COPY_7000_9500 (1-2501) x ALU6_HUMAN (1-593)

Qy 642 GTCACCCAGGATGAGTGGTGGCAACCATAGGTCACTGAGCTCAACTCTCTGAG 701

Db 308 ValalaglnAlaGlyValGlnTrpArgAspLeuGlySerProGlnProProProGly 327

Qy 702 CTCAGGAGATGCTGCTGACCTCCAGCTCCCAAGTGGGACTACGAGCGTGCACCA 761

Db 328 PhelysArgPheSerCysLeuSerLeuProSerSerTrpAspTyArgHisAlaProPro 347

Qy 762 CGCTGGCTAATTAATAAATTTTGTGACAGACTGGCTTACTAGTTGGCCAGGCT 821

Db 348 ArgProAlaAan--PheCysIlePheSerArgAspGlyValSerProCysTrpSerGly 367

Qy 822 TGTCTTAACTCTCGGCTTCAAGCAATCTCTACCTTGGCATCCCAAGTGTGGGATT 881

Db 367 rpSerArgThrProAspLeuArg**SerAlaArgLeuGlyLeuProLysCysTrpAsp 387

Qy 882 ACAGGGGTGAGCCACCATGTGGGCTACTTATTTCTTTTACATTCATCTTTCCAAAGAA 941

Db 387 yrArgArgGluProProArgProAla*****Phe----- 399

Qy 942 TGTAAAGTCCACAGACAGGAGATTACTGCCTATTTCTTCTTTTGTGACACAG 1001

Db 400 -----Phe***AspGlyV 404

Qy 1002 TCTCACTTCATCACTCAACCTCCGCT-----TCAGTCACTGCAACCTCT 1046

Db 404 alSerLeuLeuProArgLeuGluCysAsnGlyAlaIleSerAlaHisArgAsnLeuA 424

Qy 1047 GCTCCCGGGTTCAAGYGATTCCTCTGCTTAAAGCTCTCTGAGTAGCTGGAATTACAAGCG 1106

Db 424 rgLeuProGlySerSerAspSerProAlaSerAlaSerArgValAlaGlyIleThrGlyM 444

Qy 1107 TGCACACCATGCTTGGCTAAATTTTGTATTTTATAGCAGAGATGGGTTTACCATGTT 1166

Db 444 etArgHisAlaArgLeuIle-PheValPheLeuValGluThrGlyPheLeuHisVal 463

Qy 1167 GCCCAGGCTGCTCAAACTCTCTGACCTCAAGTGTATCTGCTGCCTCAGTCTCCCAAGT 1226

Db 464 GlyGlnAlaGlyLeuGluLeuProThrSerGlyAspProAlaSerAlaSerGlnSer 483

Qy 1227 GCTGGAATATAGCGCTGAGTCACTGTGCC 1256

Db 484 AlaGlyIleThrGlyValSerHisArgAla 493

RESULT 6

ALU6_HUMAN

ID ALU6_HUMAN STANDARD; PRT; 591 AA.

AC P39195;

DT 01-FEB-1995 (Rel. 31, Created)

DT 01-FEB-1995 (Rel. 31, Last sequence update)

DT 10-MAY-2005 (Rel. 47, Last annotation update)

DE Alu subfamily SX sequence contamination warning entry.

OS Homo sapiens (Human).

OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

OC Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini; Hominidae;

OC Homo.

OX NCBI_TaxID=9606;

RN [1]

RP NUCLEOTIDE SEQUENCE.

RX MEDLINE=95021758; PubMed=7935834; DOI=10.1038/371752a0;

RA Claverie J.-M., Makalowski W.;

RT "Alu alert.";

RL Nature 371:752-752 (1994).

RN [2]

RP CONCEPT.

RX MEDLINE=92241891; PubMed=1572661;

RA Claverie J.-M.;

RT "Identifying coding exons by similarity search: alu-derived and other

RT potentially misleading protein sequences.";

RL Genomics 12:838-841 (1992).

RN [3]

RP ALU FAMILIES CLASSIFICATION.

RX MEDLINE=88333009; PubMed=3138422;

RA Quentin Y.;

RT "The Alu family developed through successive waves of fixation closely

RT connected with primates lineage history.";

RL J. Mol. Evol. 27:194-202 (1988).

RN [4]

RP ALU FAMILIES CLASSIFICATION.

RX MEDLINE=91178815; PubMed=1706781;

RA Jurka J., Milosavljevic A.;

RT "Reconstruction and analysis of human Alu genes.";

RL J. Mol. Evol. 32:105-121 (1991).

CC -!- MISCELLANEOUS: Various analyses indicate that Alu repeats fall

CC into 8 subfamilies (Ref.3 and Ref.4). Therefore, 8 Alu warning

CC consensus sequences have been constituted that contain all six

CC frames conceptual translations of each of these classes of Alu

CC repeats.

CC -!- MISCELLANEOUS: Isolated 'X' indicates the presence of a stop

CC codon, 'XXX' is used to separate the various translation phases.

CC -!- CAUTION: This Alu entry is provided in order to avoid the further

CC pollution of protein sequence databases with Alu-derived amino

CC acid sequences.

CC -!- CAUTION: Alu repetitive sequences are interspersed in human and

CC primate genomes with an average spacing of 4 kb. Some of them are

CC actively transcribed by pol III. Normal transcripts may contain

CC Alu-derived sequences in 5' or 3' untranslated regions. However,

CC cDNA libraries also contain partial and/or rearranged cDNAs

ligated with Alu-derived sequence in any orientation. Although Alu elements (especially situated on the complementary strand) have a great potential to create additional/alternative exons, consideration should be given to the possibility that the presence of an Alu in an open reading frame may have resulted from a cloning artifact or may be due to misinterpretation of sequencing data. This point has been overlooked on several occasions, with the consequence of erroneous Alu-derived amino acid sequences being reported.

-!- CAUTION: Any significant similarity of a putative protein sequence with an Alu-translated entry must be taken as a warning that a part of Alu repeat may have been artifactually included in the coding nucleotide sequence.

This Swiss-Prot entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use as long as its content is in no way modified and this statement is not removed.

DR EMBL; U14574; -, NOT ANNOTATED_CDS; Genomic DNA.

FT REGION 1 96 Frame-1.
FT REGION 100 195 Frame-2.
FT REGION 199 294 Frame-3.
FT REGION 298 393 Frame-4.
FT REGION 397 492 Frame-5.
FT REGION 496 591 Frame-6.
SQ SEQUENCE 591 AA; 64434 MW; AC8154AD8A6BB280 CRC64;

Alignment Scores:

Pred. No.: 6,22e-36 Length: 591
Score: 442.00 Matches: 172
Percent Similarity: 43.2% Conservative: 37
Best Local Similarity: 35.5% Mismatches: 145
Query Match: 10.0% Indels: 130
DB: 1 Gaps: 17

US-10-607-806-1-G7328-T9182_COPY_7000_9500 (1-2501) x ALU8_HUMAN (1-591)

QY 2 TCCTGTGTCACGTGCTCCAGCTGGTAAAGAGCAAC---TCGTCTCAAAAAA----- 53
DB 178 SerArgHisCysThrProAlaThrPalaThrGlu-ArgAspSerValSerIlysIys**** 197
QY 54 -----AAAAATGCTTTCAATAATA 73
DB 197 ****ArgAlaGluAlaGlyArgGlySerArgLeu***SerGlnHisPheGlyArgly 217
QY 74 TATGATAAAGGAGCTATATATTTTCAAGCATAGGATCATTTCTCTGAGCATCTT-- 131
DB 217 slieThr***GlyGlnGluPheGluThrSerLeuAlaAsnMetValIysProArgLeuty 237
QY 132 -----GGCAGTCACTCCCACTCTGCTCTCTCTGAGAGTGGCA-- 167
DB 237 r***LysTyrLysAsn***ProGlyValValAlaArgAlaCysAsnProSerTyrSerG1 257
QY 168 -GCTGAGGCTGACCTATTCTCTGACCTTACTCTCTCTCTCTCTCTCTCTCTCTCTCT 226
DB 257 yGly***GlyArgGlyAla-----***ThrArgGluAlaGluValAlaValse 274
QY 227 CAGGTGCTGCCAGACATGACAACTCTAYGACCGCCAGCAAGCTGACAGCTGTAA 286
DB 274 rArg-----AspArgAlaThrAlaLeuGlnProGlyAr 285
QY 287 A-----TTTCTCTGGACMMCCGTACAC 310
DB 285 gGlnSerGluThrProSerGlnIlysLys*****PhePheLeuArgArgSerLeuAl 305
QY 311 CCACACTTATTCATCTCTGCTCTCTGCTGCG---GCAATCACCCTGAGCAGTGGTTAT 367
DB 305 aleuSerProArgLeuGluCysSerGlyAlaIleSerAlaHisCysAsnLeuArgLeu-- 324
QY 368 CCCTCTCTTGACCTATGAATTTCTAGTTGGTTCTCTAGTAGCGCGGGGAAATAATAGTAA 427

DB 325 -----ProGlySerSerAspSerPr 331
QY 428 CAACAGC-----CATGATTAGTGTATTTT 454
DB 331 oAlaSerAlaSerArgValAlaGlyIleThrGlyAlaArgHisAlaArgLeuIlePh 351
QY 455 CTGGTCTCTGGGAGTGTCTCTTAATCTCTCAGAACACACTATGGAGTAGTACAATT 514
DB 351 eValPheLeuValGluThrGlyPhe-----HisHisValGlyGlnAlaGlyLe 367
QY 515 ATCCTCACTTAACAGATAGAAAACAGAGCTCAGAGCTCAGAGCTATTTTGGCCAGAGAT 574
DB 367 uGluLeuLeu-----ThrSerGly-AspLeu-ProProLysVal 379
QY 575 ACACAGCTT-----GTAAGTGTGACAGTTTGGTGTGTGTGTGTGTGTGTGTGTGT 625
DB 380 LeuGlyLeuGlnAla***AlaThrAlaProGlyLeuArgPro*****PhePhe*** 399
QY 626 GACAGGCTCTGCTCTCTCACCAGGATGAGCAGATGTCGCAACCATAGGTCACTGCA 685
DB 400 AspGlyValSerLeuCysArgProGlyTyrPsrAlaValAlaArgSerArgLeuThrAla 419
QY 686 GCTCAACCTCTGAGCTCAAGGATCTCTGACCTCAGCTCCCAAGTAGTGGAGCTA 745
DB 420 ThrSerAlaSerArgValGlnAlaIleLeuLeuProGlnProGlu***LeuGlyLeu 439
QY 746 CGAGCTGACACACACAGGCTGCTGCTTAATTAATAAATTTTGTGTAGAGCTGGTCTTA 805
DB 440 GlnAlaArgAlaThrThrProGly***PheLeuTyrPhe*****ArgArgGlyPheT 459
QY 806 CTACGTTGCCAGGCTTGTCTTAACCTCTCTGCTTCAAGCAATCTCTACCTTGCATC 865
DB 459 hrMetLeuAlaArgLeuValSerAsnSer***ProGlnValIlePheLeu-----P 476
QY 866 CCAAGTGTGGGATTACAGGGGTGAGCCACCATGTCGGCTACTTATTTCTTTACATTC 925
DB 476 ro-LysCysTyrAspTyrArgArgGluProProArgProAlaSer-AlaArg***** 495
QY 926 CATCTTCCAAATAGATGTAAGATCCACAGACAGGATTAAGTCTGCTTCTCTCTCT 985
DB 495 *----- 495
QY 986 CTTTTTTGAGACAGAGCTCTCACTTCATCATCACTCAACCTCGTTTCAGC----- 1032
DB 496 -PhePheGluThrGluSerArgSerValAlaGlnAlaGlyVal-GlnTyrArgAspLeuG 515
QY 1033 --TCACTGCAACCTCTGCTCCCGGTTCAAGYGATTTCTCTGCTCAAGCTCTCTGAT 1090
DB 515 lySerLeuGlnProProProProGlyPheLysArgPheSerCysLeuSerLeuProSerS 535
QY 1091 GCTGAATTAACAGGTGACACACCATGCTTGGCTTAATTTTGTATTTTGTAGCAGAT 1150
DB 535 exTrpAspTyrArgArgAlaProProArgProAlaAsn-PheCysIlePheSerArgAsp 554
QY 1151 GGGGTTTTACCATTGTCCTGCTGCTCAAACTCTCTGACCTCAAGTATCTGCTGCTG 1210
DB 555 GlyValSerProCysTyrProGlyTyrPsrArgThrProAspLeuArg***Ser----- 572
QY 1211 CTCAGTCTCCAAAGTGTGGAAATATAGCGGTGAGTCTACTGTGCTGCGCG 1262
DB 573 ---Ser-SerGlnSerAlaGlyIleThrGlyValSerHisArgAlaAatgPro 588

RESULT 7

ALU8_HUMAN
ID: ALU8_HUMAN STANDARD; PRT; 591 AA.
AC P39195;
DT 01-FEB-1995 (Rel. 31, Created)
DT 01-FEB-1995 (Rel. 31, Last sequence update)
DT 10-MAY-2005 (Rel. 47, Last annotation update)
DE Alu subfamily SX sequence contamination warning entry.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

Db 274 rArgAspArgAlaThr-----AlaLeuGlnProGlyArgGlnSerGluThrProse 291
 QY 461 AACCAAGAAATTAACTAACTAATCATGCTGCTGTTACTATTATTTCCCGCGCCTACT 402
 Db 291 rGlnLysLys----- 294
 QY 401 GAGAACCAACTAGAAATTCATAGTCAAGGAAGGATAAACCTACTGCTACAGGTGATTGC 342
 Db 294 ----- 294
 QY 341 CGAGCCAGACGACGATGATGATAGTGTGGGTGTACGGKGTCTCCAGCAGAAATTTACA 282
 Db 295 -----PhePh 299
 QY 281 GCTGTCCAGCTTCTTGGCTGCTGTAGCAGTGTCTGTCGTCGACACCTGGAAAG 222
 Db 299 eLeuArgSerLeuAlaLeuSerProArgLeuGluCys-----Se 313
 QY 221 TGGGAGGAGCAGCTGAGATAGGATAGTGC-----AGAGCAATAGT---CAGCCCTC 171
 Db 313 rGlyAlaIleSer-----AlaHisCysAsnLeuArgLeuProGlySerSerAspSe 330
 QY 170 ACTGTCCTCCTCTCAGGAACAGGTGGCGATGACTTCCCAAGATGCTTCAGGAGA----- 116
 Db 330 rProAlaSerAlaSerArgValAlaGlyIleThrGlyAlaArgHisAlaArgLeuI 350
 QY 115 -----AATGATCTCTAT 105
 Db 350 ePheValPheLeuValGluThrGlyPheHisValGlyGlnAlaGlyLeuGluLeu 370
 QY 104 GGCTTGAAAAATATAGTCTCTTATCATATATTATTG----- 65
 Db 370 uThrSerGlyAspLeuProProlysValLeuGlyLeuGlnAla***AlaThrAlaProGI 390
 QY 64 -----AAGCATTTTTTTTTTTTGAGACAGAGTGTCTGTGTACCCAGCTGGAC 15
 Db 390 yLeuArgPro*****PhePhe***AspGlyValSerLeuCysArgProGlyTrpSe 410
 QY 14 AGCAGTGCACACGA 2
 Db 410 rAlaValAlaArg 414

RESULT 8
 ALU6_HUMAN STANDARD; PRT; 593 AA.
 AC P39193;
 DT 01-FEB-1995 (Rel. 31, Created)
 DT 01-FEB-1995 (Rel. 31, Last sequence update)
 DT 10-MAY-2005 (Rel. 47, Last annotation update)
 DE Alu subfamily SP sequence contamination warning entry.
 OS Homo sapiens (Human)
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini; Hominidae;
 OC Homo.
 OX NCBI_TaxID=9606;
 RN [1]
 RN NUCLEOTIDE SEQUENCE
 RP MEDLINE=95021758; PubMed=7935834; DOI=10.1038/371752a0;
 RA Claverie J.-M.; Makalowski W.;
 RT "Alu alert";
 RL Nature 371:752-752 (1994).
 RN [2]
 RN CONCEPT.
 RP MEDLINE=92241891; PubMed=1572661;
 RA Claverie J.-M.;
 RT "Identifying coding exons by similarity search: alu-derived and other
 RT potentially misleading protein sequences";
 RL Genomics 12:838-841 (1992).
 RN [3]
 RP ALU FAMILIES CLASSIFICATION.
 RX MEDLINE=88333009; PubMed=3138422;
 RA Quentin Y.;
 RT "The Alu family developed through successive waves of fixation closely

connected with primate lineage history."; J. Mol. Evol. 27:194-202 (1988).
 [4]
 RP ALU FAMILIES CLASSIFICATION.
 RX MEDLINE=91178815; PubMed=1706781;
 RA Jurka J., Milosavljevic A.;
 RT "Reconstruction and analysis of human Alu genes."; J. Mol. Evol. 32:105-121 (1991).
 CC -i- MISCELLANEOUS: Various analyses indicate that Alu repeats fall into 8 subfamilies (Ref.3 and Ref.4). Therefore, 8 Alu warning consensus sequences have been constituted that contain all six frames conceptual translations of each of these classes of Alu repeats.
 CC -i- MISCELLANEOUS: Isolated 'X' indicates the presence of a stop codon, 'XX' is used to separate the various translation phases.
 CC -i- CAUTION: This Alu entry is provided in order to avoid the further pollution of protein sequence databases with Alu-derived amino acid sequences.
 CC -i- CAUTION: Alu repetitive sequences are interspersed in human and primate genomes with an average spacing of 4 kb. Some of them are actively transcribed by pol III. Normal transcripts may contain Alu-derived sequences in 5' or 3' untranslated regions. However, cDNA libraries also contain partial and/or rearranged cDNAs ligated with Alu-derived sequence in any orientation. Although Alu elements (especially situated on the complementary strand) have a great potential to create additional/alternative exons, consideration should be given to the possibility that the presence of an Alu in an open reading frame may have resulted from a cloning artifact or may be due to misinterpretation of sequencing data. This point has been overlooked on several occasions, with the consequence of erroneous Alu-derived amino acid sequences being reported.
 CC -i- CAUTION: Any significant similarity of a putative protein sequence with an Alu-translated entry must be taken as a warning that a part of Alu repeat may have been artifactually included in the coding nucleotide sequence.
 CC -----
 CC This Swiss-Prot entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL Outstation - the European Bioinformatics Institute. There are no restrictions on its use as long as its content is in no way modified and this statement is not removed.
 CC -----
 CC EMBL; U14572; ; NOT ANNOTATED_CDS; Genomic_DNA.
 DR InterPro: IPR002218; GIDA.
 DR ProDom: PD003738; GIDA; 1.
 KW Hypothetical protein.
 FT REGION 1 97 Frame-1.
 FT REGION 101 196 Frame-2.
 FT REGION 200 295 Frame-3.
 FT REGION 299 395 Frame-4.
 FT REGION 399 494 Frame-5.
 FT REGION 498 593 Frame-6.
 SQ SEQUENCE 593 AA; 64640 MW; 136EF344AAACD12A2 CRC64;
 Alignment Scores:
 Pred. No.: 3,15e-34 Length: 593
 Score: 425.50 Matches: 164
 Percent Similarity: 40.0% Conservative: 36
 Best Local Similarity: 32.8% Mismatches: 138
 Query Match: 9.6% Indels: 162
 DB: 1 Gaps: 18
 US-10-607-806-1-G7328-T9182_COPY_7000_9500 (1-2501) x ALU6_HUMAN (1-593)
 QY 1260 GCCAGCAGAGTACTCAGCCCTATAATTCACGACCTTTGGGAGACTGAGCGAGGAGAT 1201
 Db 101 AlaGlyArgGlyGlySerArgLeu**SerGlnHisPheGlyArgProArgAlaAsp 120
 QY 1200 CACTTGAGTCTAGGATTTGAGACGACGCTGGGCAACATGGTAAACCCCATCTCTGCTA 1141
 Db 121 HisLeuArgSerGlyValArgAspGlnProAspGlnHisGlyGluThrProSerLeuLeu 140

Db 194 ilePheillePheillePheAsnPheLeuArgGlnSerLeuAsnSerValThrGlnAlaGly 213
 Qy 657 GCACAGTGGTGCACCAATAGGTCACTCAGCCTCAACTCTGAGCTCAAGGGATCTGCT 716
 Db 214 ValGlnTrpArgAenLeuGlySerLeuGlnProLeuProGlyPheLysLeuPheSer 233
 Qy 717 GACCTCAGCTCCCAAGTACGTGGAGTACGAGGGTGCACACCGCTGGCTTAATTA 776
 Db 234 CysProSerLeuLeuSerTrpAspTyrArgArg---ProProArgLeuAlaAsnPhe 252
 Qy 777 AAAAAATTTTTTGTAGAGACTGGGTCTTACTAGCTTGGCCAGGCTTCTTAAACTCTG 836
 Db 253 PheValPheLeuValGlu-MetGlyPheThrMetPheAlaArgLeuLeuLeuSer-- 271
 Qy 837 GCTTCAAGCAATCTCTCACTCTGGCATCCCAAGTCTGGGATTACAGGGGTGAGCCAC 896
 Db 272 GlyProCysAspLeuProAlaSerAlaSerGlnSerAlaGlylleThrGlyValSerHis 291
 Qy 897 CATGTGGGGTACTATTATTTCTTACATTCCTATCTTTCCATAGATGTAAGATCCACAGA 956
 Db 292 HisAlaArgLeuIlePheAsnPhe-----Cys----- 300
 Qy 957 ACAGGGATTACTGCTATTCTTCTCTCTCTTTTGTGAGACAGAGTCTCACTTCATCAC 1016
 Db 301 -----LeuPheGluMetGluSerHisSerValThr 310
 Qy 1017 TCAA-----CCTCGTTTCACTCACTCACTCTGCTCCCGGGTTCAA 1061
 Db 311 -GlnAlaGlyValGlnTrpProAsnLeuGlySerLeuGlnProLeuProGlyLeuLys 330
 Qy 1062 GYGATTTCTCTGCTTAAGCTCTGAGTACGTGGAATTACAGCGTGCACACCATGCTT 1121
 Db 330 ArgPheSerCysLeuSerLeuProSerSerTrpAspTyrGlyHisLeuProHisPr 350
 Qy 1122 GGCTAAATTTTGTATTTTAGCAGAGATGGGGTTTACCATGTTGCCAGGCTGGTCTC 1181
 Db 350 AlaAsn-PheCysIlePheIleArgGlyValSerProTyrLeuSerGlyTrpSerG 370
 Qy 1182 AAATCTCTGACCTCAAG 1198
 Db 370 InThrProAspLeuArg 375

RESULT 11
 Q6ZWA9_HUMAN
 ID Q6ZWA9_HUMAN PRELIMINARY; PRT; 239 AA.
 AC Q6ZWA9;
 DT 05-JUL-2004 (TrEMBLrel. 27, Created)
 DT 05-JUL-2004 (TrEMBLrel. 27, Last sequence update)
 DT 05-JUL-2004 (TrEMBLrel. 27, Last annotation update)
 DE Hypothetical protein FLJ41367.
 OS Homo sapiens (Human).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini; Hominidae;
 OC Homo.
 OX NCBI_TaxID=9606;
 RN [1]
 RP NUCLEOTIDE SEQUENCE.
 RC TISSUE=Caudate nucleus;
 RA Tanigami A., Fujiwara T., Shibahara T., Goto Y., Hirao M., Shimizu F.,
 RA Wakebe H., Ono T., Hishigaki H., Watanabe T., Ozaki K., Sugiyama T.,
 RA Irie R., Otsuki T., Sato H., Wakamatsu A., Ishii S., Yamamoto J.,
 RA Isono Y., Kawai-Hio Y., Saito K., Nishikawa T., Kimura K.,
 RA Yamashita H., Matsuo K., Nakamura Y., Sekine M., Kikuchi H., Kanda K.,
 RA Wagatsuma M., Murakawa K., Kanehori K., Takahashi-Fujii A., Oshima A.,
 RA Sugiyama A., Kawakami B., Suzuki Y., Sugano S., Nagahari K.,
 RA Maguho Y., Nagai K., Isoqai T.;
 RL Submitted (JUL-2003) to the EMBL/GenBank/DBJ databases.
 DR EMBL; AK123361; BAC05595.1; -; mRNA.
 SQ SEQUENCE 239 AA; 25635 MW; 409CED42F0603D85 CRC64;

Alignment Scores:
 Pred. No.: 2,37e-31 Length: 239
 Score: 396.50 Matches: 110

Percent Similarity: 56.0% Conservative: 25
 Best Local Similarity: 45.6% Mismatches: 68
 Query Match: 9.0% Indels: 38
 DB: 2 Gaps: 4
 US-10-607-806-1-G7328-T9182_COPY_7000_9500 (1-2501) x Q6ZWA9_HUMAN (1-239)
 Qy 625 AGACAGGTTCTGCTCTGTCTCACCCAGGCATGACACAGTGGTGCACCATAGTCACTGC 684
 Db 2 ArgTrpSerLeuThrLeuSerProArgLeuGluCysSerGlyAlaIleSerThrAspSer 21
 Qy 685 AGCTTCAACCTCTCTGAGCTCAAGGGATCTGCTGACCTTCAGCTCCCAAGTAGTGGGACT 744
 Db 22 SerLeuHisLeuProGlySerSerAsnSerLeuThrSerAlaSerGlnValGlyIle 41
 Qy 745 ACAGCGTGCACACACACCGCTGGCTTAATAATAAAAAATTTTTTTAGAGACTGGGCTT 804
 Db 42 ThrGlyAlaCysHisHisAlaArgLeuIlePheVal--PheSerValGluThrGlyPheH 61
 Qy 805 ACTAGCTGGCCAGGCTTGTCTTAACTCTGCTTCAAGCAATCTCTACCTTGGCAT 864
 Db 61 IHisValGlyGlnAlaGlyLeuGluLeuLeuThrSerAsnAspLeuProThrSerAlaS 81
 Qy 865 CCAAAAGTGTGGGATTACAGGGGTGAGCCACCATGTGGGCTACTTATTTCTTTACATT 924
 Db 81 erGlnAsnAlaGlylleThrGlyValSerHisArgAlaGln----- 94
 Qy 925 CCATCTTTTCCAATAGAAATGTAAGATCCACAGAACAGGATTAAGTCTGCTTATTTCTTCTT 984
 Db 95 -----ProGluLeuCys-----P 99
 Qy 985 TCTTTTGTGAGCAGAGTCTCACTTCACTCACTCA----- 1019
 Db 99 heileLeuLysSerPheSerPhePheLeuArgGlnGlyLeuCysHisProGlyTrpSerA 119
 Qy 1020 -----ACCTCCGTTCAGCTCAGCTGCAACCTCTGCTCCCGGGTCAAGYGATTTCTCTG 1074
 Db 119 laValAlaGlnSerGlnLeuThrGluAlaSerValSerGlnAlaGlnAlaIleLeuLeuP 139
 Qy 1075 CTAAGCTCTCTGAGTAGTGGAAATTACAGCGTGCACCATGCTGGCTTAATTTTGT 1134
 Db 139 roGlnProProGlu--TyrGlySerThrGlyMetTyrHisHisAlaTrpLeuIlePheTr 158
 Qy 1135 TATTT---TTAGCAGAGATGGGGTTTACCATGTGCCAGGCTGGTCTCAAACTCTCTGA 1191
 Db 158 p-GlyGlyLeuValAsnLysGlySerHisTyrIleAlaGlnAlaGlyLeuGluLeuLeuA 178
 Qy 1192 CCTCAAGTGATCTGCTGCTCAGTCTCTCCAAAGTGTGGAATTATAGCGTGAAGTCACT 1251
 Db 178 snSerSerAsnProProAlaSerValSerGlnLysAlaGlyIleIleAspIleSerHisC 198
 Qy 1252 GTCCGTGGCGGATGATGCTATTTTCTTATTTTCTTATTTATATCCCGAGATCTAGA 1304
 Db 198 ysAlaGlnProIleLeuGluSerLeuMetPheIleProAsnProLysGlyArg 215
 RESULT 12
 Q6ZS15_HUMAN
 ID Q6ZS15_HUMAN PRELIMINARY; PRT; 179 AA.
 AC Q6ZS15;
 DT 05-JUL-2004 (TrEMBLrel. 27, Created)
 DT 05-JUL-2004 (TrEMBLrel. 27, Last sequence update)
 DT 05-JUL-2004 (TrEMBLrel. 27, Last annotation update)
 DE Hypothetical protein FLJ45499.
 OS Homo sapiens (Human).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini; Hominidae;
 OC Homo.
 OX NCBI_TaxID=9606;
 RN [1]
 RP NUCLEOTIDE SEQUENCE.
 RC TISSUE=Thalamus;
 RA Ota T., Nakagawa S., Senoh A., Mizuguchi H., Inagaki H., Sugiyama T.,
 RA Irie R., Otsuki T., Sato H., Wakamatsu A., Ishii S., Yamamoto J.,

RA Isono Y., Kawai-Hio Y., Saito K., Nishikawa T., Kimura K.,
RA Yanashita H., Matsuo K., Nakamura Y., Sekine M., Kikuchi H., Kanda K.,
RA Wagatsuma H., Murakawa K., Kanehori K., Takahashi-Fujii A., Oshima A.,
RA Sugiyama A., Kawakami B., Suzuki Y., Sugano S., Nagahari K.,
RA Maeuho Y., Nagai K., Isogai T.,
RA Submitted (JUL-2003) to the EMBL/GenBank/DBJ databases.
RL EMBL; AK127408; BAC6963.1; -, mRNA.
DR EMBL; AK127408; BAC6963.1; -, mRNA.
SQ SEQUENCE 179 AA; 19118 MW; 52286CCCCA5A4458E CRC64;

Alignment Scores:
Pred. No.: 2,63e-30 Length: 179
Score: 386.00 Matches: 109
Percent Similarity: 60.2% Conservative: 15
Best Local Similarity: 52.9% Mismatches: 50
Query Match: 8.7% Indels: 33
DB: 2 Gaps: 6

US-10-607-806-1-G7328-T9182_COPY_7000_9500 (1-2501) x Q625I5_HUMAN (1-179)

QY 633 TCTTGTCTCTCACCAGGATGACAGACAGTGTGCAACCATAGTCTCAGCTGCAACCTCAA 692
DB 3 SerSerSerValSerGlnAlaGlyValGlnTrpLeuAspLeuSerLeuGlnProPro 22
QY 693 CTCTCTGAGCTCAAGGATCTGTGACCTCAGCTTCCCAAGTACCTGGAGCTAGAGCGT 752
DB 23 ProSerGlyPheLysArgPheSerCysLeuSerLeuGlnSerTrpAspTyrArgHis 42
QY 753 GCACACACCGCTGGCTTAATAAAAAATTTTTT-----GTAGAGACTGGTCT 803
DB 43 AlaProProCysProAlaGly-----PheCysLeuCysValGlnTrpGlyPhe 58
QY 804 TACTACCTTGGCCAGGCTGTCTTAAACTCTGTCTTCAAGCAATCTCTCTCTCTGCA 863
DB 59 HisHisValGlyGlnAlaSerLeuLysLeuLeuThrSerSerAspProAlaLeuAla 78
QY 864 TCCCAAGTCTGGGATTAAGGCTGAGCCACATGTGGCTTACTTATTTCTTACAT 923
DB 79 SerGlnIleAlaGlyIleThrGlyMetIleHisAlaGlnProLeuLeuPheLeu-- 97
QY 924 TCCATCTTTCCATAGATGTAAGTCCACAGACAGGATTAAGTCTTCTCTCTCTCT 983
DB 98 -----ArgGlnGlyValThr-----LeuLeuPro 105
QY 984 TTCTTTTGGAGACAGAGTCTCATCTCACTCAACCTCCGTTCA-GCTCACTGCAAC 1042
DB 106 -----ArgLeuGluCysSerSerAlaIleLeuAlaHisCysIle 118
QY 1043 CTCTGCTCCCGGTTCAAGYATTCCTGCTTAAAGCTCTCTGAGTCTGGAATTA 1102
DB 119 LeuCysLeuProGlySerSerAspSerProAlaSerAlaSerGlnValAlaGlyThrThr 138
QY 1103 AGCGTGCACACCATCTGGCTTAATTTTGTATTTTGTAGAGAGATGGGTTTACCA 1162
DB 139 GlyVal-----CysProLeuIle-PheValPheValGlnMetGlyPheHisI 155
QY 1163 TGTGTCCAGCTGGTCTCAAACTCTGACCTCAAGTATCTGCTCTGCTCTGCTCTCCCA 1222
DB 155 sValAla-ArgLeuValPheAsnSerSerProGlnValIleHisProProTrpProProL 175
QY 1223 AAGTGTGGAATTA 1236
DB 175 ysValLeuArgLeu 179

RESULT 13

Q8NAL9_HUMAN
ID Q8NAL9_HUMAN PRELIMINARY; PRT; 158 AA.
AC Q8NAL9;
DT 01-OCT-2002 (TrEMBLrel. 22, Created)
DT 01-OCT-2002 (TrEMBLrel. 22, Last sequence update)
DT 01-OCT-2002 (TrEMBLrel. 22, Last annotation update)
DE Hypothetical protein FLJ35131.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

OC Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini; Hominoidea;
OC Homo.
OX NCBI_TaxID=9606;
RN [1]
RP NUCLEOTIDE SEQUENCE.
RC TISSUE=Placenta;
RX PubMed=14702039; DOI=10.1038/ng1285;
RA Ota T., Suzuki Y., Nishikawa T., Otsuki T., Sugiyama T., Irie R.,
RA Wakamatsu A., Hayashi K., Sato H., Nagai K., Kimura K., Makita H.,
RA Sekine M., Obayashi M., Nishi T., Shibahara T., Tanaka T., Ishii S.,
RA Yamamoto J.-I., Saito K., Kawai T., Isono Y., Nakamura Y.,
RA Nagahari K., Murakami K., Yasuda T., Iwayanagi T., Wagatsuma M.,
RA Shiratori A., Sudo H., Hosoiri T., Kaku Y., Kodaira H., Kondo H.,
RA Sugawara M., Takahashi M., Kanda K., Yokoi T., Furuya T., Kikkawa E.,
RA Omura Y., Abe K., Kamihara K., Katsuta N., Sato K., Tanikawa M.,
RA Yamazaki M., Ninomiya K., Ishibashi T., Yamashita H., Murakawa K.,
RA Fujimori K., Tanai H., Kimata M., Watanabe M., Hiraoa S., Chiba Y.,
RA Ishida S., Ono Y., Takiguchi S., Watanabe S., Yosida M., Hotuta T.,
RA Kusano J., Kanehori K., Takahashi-Fujii A., Hara H., Tanase T.-O.,
RA Nomura Y., Togai S., Komai P., Hara R., Takeuchi K., Arita M.,
RA Imose N., Musashino K., Yuuki H., Oshima A., Sasaki N., Aotsuka S.,
RA Yoshikawa Y., Matsunawa H., Ichihara T., Shiohata N., Sano S.,
RA Moriya S., Momiyama H., Satoh N., Takami S., Terashima Y., Suzuki O.,
RA Nakagawa S., Senoh A., Mizoguchi H., Goto Y., Shimizu F., Wakebe H.,
RA Hishigaki H., Watanabe T., Sugiyama A., Takemoto M., Kawakami B.,
RA Yamazaki M., Watanabe K., Kumagai A., Itakura S., Fukuzumi Y.,
RA Fujimori Y., Komiyama M., Tashiro H., Tanigami A., Fujiwara T.,
RA Ono T., Yamada K., Fujii Y., Ozaki K., Hirao M., Ohmori Y.,
RA Kawabata A., Hikiji T., Kobatake N., Inagaki H., Ikema Y., Okamoto S.,
RA Okitani R., Kawakami T., Noguchi S., Itoh T., Shigeta K., Senba T.,
RA Matsumura K., Nakajima Y., Mizuno T., Morinaga M., Sasaki M.,
RA Togashi T., Oyama M., Hata H., Watanabe M., Komatsu T.,
RA Mizushima-Sugano J., Satoh T., Shirai Y., Takahashi Y., Nakagawa K.,
RA Okumura K., Negase T., Nomura N., Kikuchi H., Maeuho Y., Yamashita R.,
RA Nakai K., Yada T., Nakamura Y., Ohara O., Isogai T., Sugano S.;
RT "Complete sequencing and characterization of 21,243 full-length human
RT cDNAs.";
RL Nat. Genet. 36:40-45(2004).
DR EMBL; AK092450; BAC03893.1; -, mRNA.
DR Ensembl; ENSG00000198057; Homo sapiens.
SQ SEQUENCE 158 AA; 17414 MW; E15103015P2FE34 CRC64;

Alignment Scores:

Pred. No.: 3,62e-30 Length: 158
Score: 384.50 Matches: 96
Percent Similarity: 47.2% Conservative: 6
Best Local Similarity: 44.4% Mismatches: 44
Query Match: 8.7% Indels: 71
DB: 2 Gaps: 3

US-10-607-806-1-G7328-T9182_COPY_7000_9500 (1-2501) x Q8NAL9_HUMAN (1-158)

QY 633 TCTTGTCTCTCACCAGGATGACAGACAGTGTGCAACCATAGTCTCAGCTGCAACCTCAA 692
DB 3 SerHisSerValThrGlnAlaGlyValGlnTrpCysGlyLeuGlySerLeuGlnProLeu 22
QY 693 CTCTCTGAGCTCAAGGATCTGTGACCTCAGCTTCCCAAGTACCTGGAGCTAGAGCGT 752
DB 23 ProProGlyPheLysArgPheCysLeuSerLeuLeuSerSerTrpAspTyrArgArg 42
QY 753 GCACACACACCGCTGGCTTAATAAAAAATTTTTTGTAGAGACTGGTCTTACTAGCTT 812
DB 43 ValProProProLeuAlaAsn----- 49
QY 813 GGCAGGCTTGTCTTAAACTCTCTGGCTTCAAGCAATCTCTTACTTGGCATCCCAAGT 872
DB 50 -----PheCysIle----- 52
QY 873 GCTGGGATTACAGGGGTGAGCCACCATGTGCGGCTACTTATTCTTTTACATTCCTTT 932
DB 52 ----- 52
QY 933 CCAATAGATGTAAGATCCACAGACAGGGGATTACTGCTCTATTTCTCTCTTTCTTTT 992

```
Db 53 -----PhePheSerPhePhePhe 58
Qy 993 GAGACAGAGTCTACTTCATCACCCTCAACCTCCGCTCCAGC-----TCACT 1037
Db 59 GlulysGluSerLeuSerValThrGlnAlaGlyVal-GlnTrpHisAspLeuGlySerLe 78
Qy 1038 GCACCTCTGCCTCCCGGCTCAAGYGATTCCTCGCTCAAGCTCTCTGAGTAGCTGGAA 1097
Db 78 uGlnAlaAlaProProGlyPheThrProPheSerCysLeuSerLeuProSerSerTrpAs 98
Qy 1098 TTACAAGCTGCACACCATCTTGGCTAA-TTTTGTGTTATTTTACAGAGATGGGGTT 1156
Db 98 nTyArgArgProProProCysProAlaAsnPhePheValPheLeuValGluMetGlyPh 118
Qy 1157 TTACCAAGTGTCCAGCAGCTGTCTCAAACTCTGACCTCAAGTGATCTGCTGCTCAGT 1216
Db 118 eCysHisIleGlyGlnAlaGlyLeuGluLeuThrSerSerAspProProThrSerAl 138
Qy 1217 CTCCCAAGTCTGGATTTATAGCGTGAGTCACTGCTGCTGGCGG 1262
Db 138 aSerGlnSerAlaGlyIleThrGlyMetSerHisArgThrTrpPro 153
```

RESULT 14

```
ALU5_HUMAN
ID _ALU5_HUMAN STANDARD; PRT; 585 AA.
AC P39192;
DC 01-FEB-1995 (Rel. 31, Created)
DT 01-FEB-1995 (Rel. 31, Last sequence update)
DT 10-MAY-2005 (Rel. 47, Last annotation update)
DE Alu subfamily SC sequence contamination warning entry.
OS Homo sapiens (Human)
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini; Hominidae;
OC Homo.
OC NCBI_TaxID=9606;
RN [1]
RP NUCLEOTIDE SEQUENCE.
RX MEDLINE=95021758; PubMed=7935934; DOI=10.1038/371752a0;
RA Claverie J.-M., Makalowski W.;
RT "Alu alert.";
RL Nature 371:752-752(1994).
RN [2]
RP CONCEPT.
RX MEDLINE=92241891; PubMed=1572661;
RA Claverie J.-M.;
RT "Identifying coding exons by similarity search: alu-derived and other
RT potentially misleading protein sequences.";
RL Genomics 12:838-841(1992).
RN [3]
RP ALU FAMILIES CLASSIFICATION.
RX MEDLINE=88333009; PubMed=3138422;
RA Quentin Y.;
RT "The Alu family developed through successive waves of fixation closely
RT connected with primate lineage history.";
RL J. Mol. Evol. 27:194-202(1988).
RN [4]
RP ALU FAMILIES CLASSIFICATION.
RX MEDLINE=91178815; PubMed=1706781;
RA Jurka J., Milosavljevic A.;
RT "Reconstruction and analysis of human Alu genes.";
RL J. Mol. Evol. 32:105-121(1991).
CC -!- MISCELLANEOUS: Various analyses indicate that Alu repeats fall
CC into 8 subfamilies (Ref.3 and Ref.4). Therefore, 8 Alu warning
CC consensus sequences have been constituted that contain all six
CC frames conceptual translations of each of these classes of Alu
CC repeats.
CC -!- MISCELLANEOUS: Isolated 'X' indicates the presence of a stop
CC codon, 'XXX' is used to separate the various translation phases.
CC -!- CAUTION: This Alu entry is provided in order to avoid the further
CC pollution of protein sequence databases with Alu-derived amino
CC acid sequences.
CC -!- CAUTION: Alu repetitive sequences are interspersed in human and
```

```
CC primate genomes with an average spacing of 4 kb. Some of them are
CC actively transcribed by pol III. Normal transcripts may contain
CC Alu-derived sequences in 5' or 3' untranslated regions. However,
CC cDNA libraries also contain partial and/or rearranged cDNAs
CC ligated with Alu-derived sequence in any orientation. Although Alu
CC elements (especially situated on the complementary strand) have a
CC great potential to create additional/alternative exons,
CC consideration should be given to the possibility that the presence
CC of an Alu in an open reading frame may have resulted from a
CC cloning artifact or may be due to misinterpretation of sequencing
CC data. This point has been overlooked on several occasions, with
CC the consequence of erroneous Alu-derived amino acid sequences
CC being reported.
CC -!- CAUTION: Any significant similarity of a putative protein sequence
CC with an Alu-translated entry must be taken as a warning that a
CC part of Alu repeat may have been artifactually included in the
CC coding nucleotide sequence.
CC -----
CC This Swiss-Prot entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL Outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use as long as its content is in no way modified and this statement is not
CC removed.
CC -----
CC EMBL; U14571; -; NOT ANNOTATED_CDS; Genomic_DNA.
CC -----
KW Hypothetical protein_95 Frame-1.
FT REGION 1 95 Frame-2.
FT REGION 99 193 Frame-3.
FT REGION 197 291 Frame-4.
FT REGION 295 389 Frame-5.
FT REGION 393 487 Frame-6.
FT REGION 491 585
SQ SEQUENCE 585 AA; 63994 MW; 46E8C4F493650A7 CRC64;

Alignment Scores:
Pred. No.: 8.67e-30 Length: 585
Score: 382.50 Matches: 108
Percent Similarity: 56.8% Conservative: 18
Best Local Similarity: 48.6% Mismatches: 65
Query Match: 8.6% Indels: 31
DB: 1 Gaps: 2

US-10-607-806-1-G7328-T9182_COPY_7000_9500 (1-2501) x ALU5_HUMAN (1-585)

Qy 1260 GCCAGGCACAGTCACTCAGCCCTATATTCACGACCTTTGGGAGACTGGCAGCGAGAT 1201
Db 99 AlaglyArgGlyGlySerArgLeu***SerGlnHisPheGlyArgProArgGalaasp 118
Qy 1200 CACTTGAGGTCCAGAGTTTGAGACCAGCCCTGGGCAACATCGTAAACCCCATCTCTCTA 1141
Db 119 His--GluValLysArgSerArgProSerTrpProThrTrp***AsnProValSerThrL 138
Qy 1140 AAAATACAAAATAATTAGCCAGCATGGTGTGTCACGCTTGTATTCAGCTACTCAGAG 1081
Db 138 ysaAnThr-LysIleSerTrpAlaTrpTrpArgAlaProValProAlaThrArgGlu 157
Qy 1080 GCTTAGCCAGCAGATCCTTGAACCGGAGCGAGGTTCGAGTGAGTGCAGCGAGG 1021
Db 158 AlaGluAlaGlyGluSerLeuGluProGlyArgArgLeuGln***AlaGlu----- 175
Qy 1020 TTGAGGTGATGAAGTGAGACTCTGTCTCAAAAAAGGAAGAAATAGGCAGTAATCC 961
Db 176 -----Ile 176
Qy 960 CTGTTCTGTGGATCTTACATCTTATTCGAAAGATGGAATGTAAGAATAAAG-----TA 907
Db 177 AlaProLeuHisSerSerLeuAlaThrGluArgAspSerValSerLys-Lys***** 196
Qy 906 GCCGCACATGGTGCCTACCCCTGTATCCAGCAGCTTTGGGATGCCAAGTAGGAGAT 847
Db 196 *ProGlyAlaAlaHisAlaCysAsnProSerThrLeuGlyGlyArgGlyGlyArgI 216
Qy 846 TGCTTGAAGCCAGGAGTTTAAGACAAGCCTGGCCACGCTAGTAGAACCAGCCAGTCTCTACAA 787
```

```
Db 216 eThArgSerArgAspArg--AspHisProGlyGlnHisGlyGluThrProSerLeu--L 235
QY 786 AAAAAATTTTAAATGACAGGCGTGTGTGACAGCTCGTAGTCCAGCTACTTGGGA 727
Db 235 euly81leGlnLysLeuAlaGlyArgGlyGlyAlaArgLeu***SerGlnLeuLeuGlyA 255
QY 726 GGCTGAGGTGACGAGATCCCTTGAGCTCAGGAGGTGAGGCTGAGCTGAGCTATGGTTGC 667
Db 255 xGLeuArgGlnGluAsnArgLeuAsnProGlyGlyGlyGlyCysSerGluProArgSera 275
QY 666 ACACACTGTGCTATGCTGCTGCTGCTGACAGAGCAAGACCTGCTCTTAACAA 617
Db 275 rghisCysThrProAlaTrpArgGlnSerGluThrProSerGlnLysLys 291
```

RESULT 15

```
ALU3_HUMAN
ID -ALU3_HUMAN STANDARD; PRT; 587 AA.
AC P39190;
DT 01-FEB-1995 (Rel. 31, Created)
DT 01-FEB-1995 (Rel. 31, Last sequence update)
DT 10-MAY-2005 (Rel. 47, Last annotation update)
DE Alu subfamily S81 sequence contamination warning entry.
OS Homo sapiens (human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini; Hominidae;
OC Homo.
OX NCBI_TaxID=9606;
RN [1]
RP NUCLEOTIDE SEQUENCE.
RX MEDLINE=95021758; PubMed=7935834; DOI=10.1038/371752a0;
RA Claverie J.-M., Makalowski W.;
RT "Alu alert.";
RL Nature 371:752-752 (1994).
RN [2]
RP CONCEPT.
RX MEDLINE=92241891; PubMed=1572661;
RA Claverie J.-M.;
RT "Identifying coding exons by similarity search: alu-derived and other
RT potentially misleading protein sequences.";
RL Genomics 12:838-841 (1992).
RN [3]
RP ALU FAMILIES CLASSIFICATION.
RX MEDLINE=88333009; PubMed=3138422;
RA Quentin Y.;
RT "The Alu family developed through successive waves of fixation closely
RT connected with primate lineage history.";
RL J. Mol. Evol. 27:194-202 (1988).
RN [4]
RP ALU FAMILIES CLASSIFICATION.
RX MEDLINE=91178815; PubMed=1706781;
RA Jurka J., Milosavljevic A.;
RT "Reconstruction and analysis of human Alu genes.";
RL J. Mol. Evol. 32:105-121 (1991).
CC -!- MISCELLANEOUS: Various analyses indicate that Alu repeats fall
CC into 8 subfamilies (Ref.3 and Ref.4). Therefore, 8 Alu warning
CC consensus sequences have been constituted that contain all six
CC frames conceptual translations of each of these classes of Alu
CC repeats.
CC -!- MISCELLANEOUS: Isolated 'X' indicates the presence of a stop
CC codon, 'XXX' is used to separate the various translation phases.
CC -!- CAUTION: This Alu entry is provided in order to avoid the further
CC pollution of protein sequence databases with Alu-derived amino
CC acid sequences.
CC -!- CAUTION: Alu repetitive sequences are interspersed in human and
CC primate genomes with an average spacing of 4 kb. Some of them are
CC actively transcribed by pol III. Normal transcripts may contain
CC Alu-derived sequences in 5' or 3' untranslated regions. However,
CC cDNA libraries also contain partial and/or rearranged cDNAs
CC ligated with Alu-derived sequence in any orientation. Although Alu
CC elements (especially situated on the complementary strand) have a
CC great potential to create additional/alternative exons,
CC consideration should be given to the possibility that the presence
```

```
CC of an Alu in an open reading frame may have resulted from a
CC cloning artifact or may be due to misinterpretation of sequencing
CC data. This point has been overlooked on several occasions, with
CC the consequence of erroneous Alu-derived amino acid sequences
CC being reported.
CC -!- CAUTION: Any significant similarity of a putative protein sequence
CC with an Alu-translated entry must be taken as a warning that a
CC part of Alu repeat may have been artifactually included in the
CC coding nucleotide sequence.
CC -----
CC This Swiss-Prot entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL Outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use as long as its content is in no way modified and this statement is not
CC removed.
CC -----
CC EMBL; U14569; -; NOT_ANNOTATED_CDS; Genomic_DNA.
KW Hypothetical protein.
FT REGION 1 96 Frame-1.
FT REGION 100 194 Frame-2.
FT REGION 198 292 Frame-3.
FT REGION 296 391 Frame-4.
FT REGION 395 489 Frame-5.
FT REGION 493 587 Frame-6.
SQ SEQUENCE 587 AA; 63613 MW; 85C4155726DEF235 CRC64;
Alignment Scores:
Pred. No.: 8.68e-30 Length: 587
Score: 382.50 Matches: 141
Percent Similarity: 38.3% Conservative: 24
Best Local Similarity: 32.7% Mismatches: 140
Query Match: 8.6% Indels: 126
DB: 1 Gaps: 11
US-10-607-806-1-G7328-T9182_COPY_7000_9500 (1-2501) x ALU3_HUMAN (1-587)
QY 2 TCGTGTCACTGTCTCCAGCTGGGTAAACAGAGCAACTCTGTCTCAAAAAAATAATGC 61
Db 275 SerArgHisCysThrProAlaTrpAlaThrGluArgAspSerValSerLysLys----- 292
QY 62 TTTCAATAAATATATGATAAAGGACTTATATTTTCAAGCCATAGGATCATTTCTCCT 121
Db 293 -----*****PhePheLeuArgArgSerLeuAlaLeuSerPro 306
QY 122 -----GRAGCATCTTGGCGAAGTCAATCCCACTCTTCTGAGAGTGGCAGGTGAGG 175
Db 307 ArgLeuGluCysSerGlyGlyIleSerAlaHisCysLysLeuArgLeuProGly----- 324
QY 176 CTGACCTATTGCTGTCACCTTACTCTCTATCTCAGCTGTCCCTCCCACTTTCCAGGTGCTG 235
Db 325 -----SerArgHisSerProAlaSerAla----- 332
QY 236 CCAGACACATGACAACTGCTAYGACCCAGCCAGAGTGGACAGCTGTAAATTTCTGCT 295
Db 333 -----SerGlnValaLalaLala-----ThrThr 339
QY 296 GGACAMMCCGTACACCCACACCTA---TTCACTACTGCTGCTCTGGCTCGGCAATCACCTG 352
Db 340 GylalaArgHisTyraAlaArgLeuIlePheValPheLeuValGluThrGlyPheHis--- 358
QY 353 TAGCAGTAGGTTTATCCCTTCTCTTGACCTATGAATTTCTAGTTGTTCTCAGTAGGCCGG 412
Db 358 ----- 358
QY 413 GGGAAATATAGTAACAACAGCCATGATTAGTTGTTTAAATTTTCTGGGCAAGTGT 472
Db 359 -----ArgPheSerArgAspGlyLeuAsp 366
QY 473 CTCCTTTAATCTCTCAGAACACACATATGGATAGGTACAAATTATCTCTCACTTAACAGATA 532
Db 367 LeuLeuThrSer***SerAlaArgLeuGly----- 376
QY 533 AGAAAACTGAGGCTCAGAGGCTGAGCTATTATTTGCCCAAGATCACACAGCTTGTAACTGTT 592
```

```
Db 377 -----|||||
Qy 593 GACAGTTGGGTTTTTTTGTGTTAGAGACAGGGTCTGCTCTGTCACCCAGGC 652
Db 387 ProProArgProAla*****PhePhe***AspGlyValSerLeuCysArgProGly 406
Qy 653 ATGAGCACAGTGGTGCAACCATAGTCACTGCAGCCTCAACCTCTGAGCTCAAGGGATC 712
Db 407 TrpSerAlaValAlaGlySerArgLeuThrAlaSerSerAlaSerArgValHisAlaIle 426
Qy 713 TGCTGACCTCAGCCTCCCAAGTAGCTGGGACTACGAGCGTGCCACCCAGCCTGGCTAA 772
Db 427 LeuLeuProGlnProProLys**LeuGlyLeuGlnAlaProAlaThrThrProGly** 446
Qy 773 TTAATAAAATTTTTTTGTAGAGACTGGGTCTTACTAGTTGGCCAGGCTGTCTTAACT 832
Db 447 PheLeuTyrPhe*****Arg-ArgGly---PheThrValLeuAlaGlyMetValSerI 465
Qy 833 CCTGGCTTCAAGCAATCCTCTACCTTGGCATCCCAAGTCTGGGATTACAGGGGTGAG 892
Db 465 eSer***ProArgAspProProAlaSerAlaSerGlnSerAlaGlyIleThrGlyValSe 485
Qy 893 CCACCATGTGGGCTACTTATTTCTTTACATTCATCTTTCCAATAGATGAATGATCCA 952
Db 485 rHisArgAlaArg*****----- 491
Qy 953 CAGAACAGGGATTACTGCCTATTTCTTCTTTTGTGAGACAGAGTCTACCTTCAT 1012
Db 492 -----***PheGluThrGluSerArgSerVa 501
Qy 1013 CACCTCAACCTCCGTTTCAGC-----TCACCTGCACCTCTGCCTCCCGGT 1057
Db 501 lAlaGlnAlaGlyVal-GlnTrpArgAspLeuGlySerLeuGlnAlaProProGlyP 521
Qy 1058 TCAAGYGATTCTCCTGCTAAGCCTCTGAGTACTGGAATTACAAGCGTGCCACCACAT 1117
Db 521 heThrProPheSerCysLeuSerLeuProSerSerTrpAspTyrArgArgProProLeuA 541
Qy 1118 GCTTGGCTAATTTTTTTGTATTTTGTAGAGATGGGGTTTTTACCATGTGCCAGGCTGG 1177
Db 541 rgProAlaAsn-PheCysIlePheSerArgAspGlyValSerProPhe***ProGlyTrp 560
Qy 1178 TCTCAAACTCCTGACCTCAAGTATCTGCCTGCCTCAGTCTCCCAAAGTGTGGAATTAT 1237
Db 561 SerArgSerProAspLeu--ValIleArgProProArgProProLysValLeuGlyLeuG 580
Qy 1238 AGCGTGAGTCATGTGCTGGC 1260
Db 580 lnAla***AlaThrAlaProGly 587
```

Search completed: February 8, 2006, 11:14:43
Job time : 596.638 secs

GenCore version 5.1.7
Copyright (c) 1993 - 2006 Bioacceleration Ltd.

OM nucleic - protein search, using frame_plus_n2p model

Run on: February 8, 2006, 09:18:47 ; Search time 16.3384 Seconds
(without alignments)
2694.617 Million cell updates/sec

Title: US-10-607-806-1_COPY_5000_5500
Perfect score: 900
Sequence: 1 acggggctactctgtgcc.....tttaataaagacggggttc 501

Scoring table:
BLOSUM62
Xgapop 10.0 , Xgapext 0.5
Ygapop 10.0 , Ygapext 0.5
Fgapop 6.0 , Fgapext 7.0
Delop 6.0 , Delext 7.0

Searched: 2443163 seqs, 439378781 residues

Total number of hits satisfying chosen parameters: 4886326

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Command line parameters:

-MODEL=frame+ n2p.model -DEV=xlp
-Q=/abss/ABSSWEB spool/US10607806/runat_08022006_091817_26823/app_query.fasta_1
-DB=A Geneseq -QFMT=fastan -SUFFIX=n2p.rag -MINMATCH=0.1 -LOOPCL=0 -LOOPEXT=0
-UNITS=bits -START=1 -END=1 -MATRIX=blosum62 -TRANS=human40.cdi -LIST=45
-DOCALIGN=200 -THR SCORE=pct -THR MAX=100 -THR MIN=0 -ALIGN=15 -MODE=LOCAL
-OUTFMT=ptc -NORM=ext -HEAPSIZE=500 -MINLEN=0 -MAXLEN=200000000 -HOST=abss02p
-USER=US10607806 @CGN 1.1 605 @runat_08022006_091817_26823 -NCPU=6 -ICPU=3
-NO_MMAP -NEG SCORES=0 -WAIT -DSPBLOCK=100 -LONGLOG -DEV TIMEOUT=120
-WARN TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5 -FGAPOP=6 -FGAPEXT=7
-YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database : A Geneseq 21.*
1: Geneseqp1980s.*
2: Geneseqp1990s.*
3: Geneseqp2000s.*
4: Geneseqp2001s.*
5: Geneseqp2002s.*
6: Geneseqp2003as.*
7: Geneseqp2003bs.*
8: Geneseqp2004s.*
9: Geneseqp2005s.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	391	43.4	170	7 ADB64987	Adb64987 Human pro
2	381.5	42.4	382	4 AAU32610	Aau32610 Novel hum
3	381.5	42.4	382	4 AAU31818	Aau31818 Novel hum
4	381.5	42.4	382	4 AAU32707	Aau32707 Novel hum
5	378	42.0	213	4 AAM25396	Aam25396 Human pro
6	373.5	41.5	186	8 AAM11829	Aam11829 Human pro
7	371.5	41.3	324	4 AAU29573	Aau29573 Novel hum
8	371.5	41.3	324	4 AAU29579	Aau29579 Novel hum
9	366.5	40.7	241	4 ABB11464	Abb11464 Human neu

10	366.5	40.7	361	5 ABG68738	Abg68738 Human pro
11	366	40.7	381	4 AAU30235	Aau30235 Novel hum
12	363	40.3	217	7 ADC87151	Adc87151 Human GPC
13	358.5	39.8	179	8 ADR09211	Adr09211 Human pro
14	355	39.4	196	7 ADE09985	Ade09985 Novel pro
15	353	39.2	296	4 AAU33304	Aau33304 Novel hum
16	345	38.3	154	4 ABG08425	Abg08425 Novel hum
17	344.5	38.3	158	7 ADB64893	Adb64893 Human pro
18	343.5	38.2	302	5 ADK36936	Adk36936 Novel hum
19	342	38.0	203	7 ADM04063	Adm04063 Human pro
20	341.5	37.9	257	4 AAO10294	Aao10294 Human pol
21	338.5	37.6	201	7 ADC86633	Adc86633 Human GPC
22	338.5	37.6	411	4 ABG08428	Abg08428 Novel hum
23	329.5	36.6	154	6 ADA54692	Ada54692 Human pro
24	328.5	36.5	293	4 AAU31676	Aau31676 Novel hum
25	328	36.4	158	4 AAU30493	Aau30493 Novel hum
26	321	35.7	218	7 ADC86939	Adc86939 Human GPC
27	316.5	35.2	185	4 AAO06014	Aao06014 Human pol
28	314.5	34.9	375	3 AAB01399	Aab01399 Neuron-as
29	314.5	34.9	375	5 ABB81538	Abb81538 Neural th
30	314.5	34.9	375	5 AAE29142	Aae29142 AD7C-NTP
31	314.5	34.9	375	6 ABR63268	AbR63268 AD7C-NTP
32	314.5	34.9	375	6 ADA84017	Ada84017 Human POM
33	314.5	34.9	375	6 ABU03024	Abu03024 Human neu
34	314.5	34.9	375	6 ABB99774	Abb99774 Amino aci
35	314.5	34.9	375	6 ABJ19445	Abj19445 AD7C-neur
36	314.5	34.9	375	7 ADB37642	Adb37642 Human neu
37	314.5	34.9	375	8 ADR14409	Adr14409 Human NF-
38	314.5	34.9	375	9 AEA30020	Aea30020 Human neu
39	312	34.6	603	7 ADJ68334	Adj68334 Human hea
40	311	34.6	239	8 ADQ65417	Adq65417 Novel hum
41	305	33.9	215	4 AAU31513	Aau31513 Novel hum
42	300	33.3	167	7 ADC86521	Adc86521 Human GPC
43	299.5	33.3	189	7 ADB65042	Adb65042 Human pro
44	298.5	33.2	290	4 ABG21913	Abg21913 Novel hum
45	298	33.1	146	4 ABG08195	Abg08195 Novel hum

ALIGNMENTS

RESULT 1

ADB64987
ID ADB64987 standard; protein; 170 AA.
XX
AC ADB64987;
XX
DT 04-DEC-2003 (first entry)
XX
DE Human protein encoded by clone PROST20121570.
XX
KW Human; pharmaceutical; diagnostic; gene therapy; tissue regeneration;
KW cell regeneration; membrane protein; signal transduction-related protein;
KW transcription-related protein; osteoporosis; neurological disease;
KW cancer; tumour.
XX
OS Homo sapiens.
XX
PN EP1308459-A2.
XX
PD 07-MAY-2003.
XX
PF 28-MAR-2002; 2002EP-00007401.
XX
PR 05-NOV-2001; 2001JP-00379298.
XX
PR 25-JAN-2002; 2002US-00350978.
XX
(HELI-) HELIX RES INST.
(REAS-) RES ASSOC BIOTECHNOLOGY.

PI Isogai T, Sugiyama T, Otsuki T, Wakamatsu A, Sato H, Ishii S;
PI Yamamoto J, Isono Y, Hio Y, Otsuka K, Nagai K, Irie R, Tamechika I;
PI Seki N, Yoshikawa T, Otsuka M, Nagahari K, Masuho Y;
XX

US-10-607-806-1 COPY 5000 5500 (1-501) x AAU32610 (1-382)

Qy	8	TCACTCTGTCGCCAGCGTGGAGTGCGGTGGCGATCTCAGATCACATACAACTCCATC	67
Db	103	SerLeuCysTyrProGlyTrpSerArgValAla**SerArgIleThrAlaThrSer-Al	1322
Qy	68	TCCTGGTTCAAATAATTCTCTTGCTCAGCGCTCTCAAGTAGCTGGGACITGTAGCTCTC	1277
Db	122	aProGlyLeuLys***PheAlaCyAPheserLeuProserSerArgAspTyr-----	139
Qy	128	AAGTAGCTGGCACACACCACCATGCCAGCTAAATTTTTGTGTCTTTTTTGGTAGAGA	187
Db	140	-----ArgHisValProProHisProGlyAsnPhe---CysIle---PheGlyALGAS	155
Qy	188	CAGGTTTTCCACCATGTGGCCAGGCTGGGTGACCTCCCTTTTAGATTCTCCTCATCTGC	247
Db	155	pGluValSerProCysTrpProGlyTrpPhe***Thr-----ProAs	169
Qy	248	TCATTCTTCCCCTTTCTAATGCAGTATCCAGTTTCCTTA-----	287
Db	169	pLeuArgTyrPro-----ProAlaSerAlaSerGlnSerAlaGluLeil	184
Qy	288	-----CTTATCACATTATTATTATTATTATTAT	316
Db	184	eGlyValSerHisHisThrTrpProGlnGluValPheLeuPheLeuAsnLeuPhelley	204
Qy	317	TATTGAGACAGAGTCTTGCTTTGTGCGCAAGCTGGAGTACAGTGGTGGCATCTCGGCTC	376
Db	204	rLeuArgrTrpSerLeuAspSerValAlaGlnAlaArgValGlnArgargAspLeuGlySe	224
Qy	377	ACTGCAAGCTCCACCTGCTGGGTTTACGCCCATTTCCCGCTCAGCTCCCCAGTAGCTG	436
Db	224	rLeuGlnAlaproProProArpPheLyPheProPheSerCysLeuSerLeuProserSerTr	244
Qy	437	GGACTAAAGCGCGCTGCCACCACGCCCGCTAAATTTTTTGTATTTTTTAATAAGACGGG	496
Db	244	pAspTyrArgAproProProHisProAlaAsnPhePheValPheLeuValGluThrGl	264
Qy	497	GTTT	500
Db	264	yPhe	265

PT Nucleic acids encoding a range of human polypeptides, useful in genetic
 PT vaccination, testing and therapy.
 XX
 PS Claim 20; Page 513; 765pp; English.

The invention relates to novel human secreted polypeptides. The polypeptides and antibodies to the polypeptides are useful for determining the presence of or predisposition to a disease associated with altered levels of polypeptide. The polypeptides are also useful for identifying agents (agonists and antagonists) that bind to them. Cells expressing the proteins are useful for identifying a therapeutic agent for use in treatment of a pathology related to aberrant expression or physiological interactions of the polypeptide. Vectors comprising the nucleic acids encoding the polypeptides and cells genetically engineered to express them are also useful for producing the proteins. The proteins are useful in genetic vaccination, testing and therapy, and can be used as nutritional supplements. They may be used to increase stem cell proliferation; to regulate haematopoiesis; and in bone, cartilage, tendon and/or nerve tissue growth or regeneration; immune suppression and/or stimulation; as anti-inflammatory agents; and in treatment of leukaemias. AAU29510-AAU33304 represent the amino acid sequences of novel human secreted proteins of the invention.

Sequence 382 AA;

Alignment Scores:	
Pred. No.:	5,39e-36
Score:	381.50
Length:	382
Matches:	90
Conservative:	13
Mismatches:	43
Indels:	36
Gaps:	6
DB:	4
Percent Similarity:	56.8%
Best Local Similarity:	49.5%
Query Match:	42.4%

US-10-607-806-1_COPY_5000_5500 (1-501) x AAU1818 (1-382)

Qy 8 TCACTGTGTCCCCAGGCTGGAGTGGCGTGGATCTCAGATCACTACACCTCCATC 67
|||
Db 103 SerLeuCystrProGlyTyrSerArgValAla**SerArgIleThrAlaThrSer-Al 122
|||
Qy 68 TCTGGGTTCAAATAATCTCTGGCTGCAGCTCTCAAGTAGCTGGAGCTTGTAAGTCTC 127
:||||| ||| :|||: ||| ||| ||| |||
Db 122 aProGlyLeuIlys**PheAlaCysPheSerLeuProSerSerArgAspTyr----- 139
|||
Qy 128 AAGTAGCTGGCACACACACACACCTCCAGCTCAATTTTGTGTGTTTTTTTGGTAGAGA 187
|||
Db 140 -----ArgHisValProHisProGlyAsnPhe---CysIle---PheGlyArgAs 155
|||
Qy 188 CAGGTTTTTACCAGTTGGCCAGCTGGGTGACCTGCCCTTTTAGATTCTCTCATCTCTGC 247
|||
Db 155 pGluValSerProCysIrrProGlyTyrPhe**Thr-----ProAs 169
|||
Qy 248 TCTATTCTCCCTTCTTAATGCAGTAGTATCCAGTTTCTTTA----- 287
|||
Db 169 pLeuArgTyrPro-----ProAlaSerAlaSerGlnSerAlaGluIleI 184
|||
Qy 288 -----CTTATCACATTATTATTATTCTTTATTAT 316
|||
Db 184 eGlyValSerHisHisThrTrpProGlnGluValPheLeuPheLeuAsnLeuPheIleTy 204
|||
Qy 317 TATTGACAGAGTCTTGCTTTGTGCGCAAGCTGGAGTAGTACAGTGGTCCGATCTCGGCTC 376
|||
Db 204 rLeuArgTrpSerLeuAspSerValAlaGlnAlaArgValGlnArgArgAspLeuGlySe 224
|||
Qy 377 ACTGCAAGCTCCACCTGCTGGTTTCACGCCATTCTCCGGCTCAGCTCCCGCAGTAGCTG 436
|||
Db 224 rLeuGlnAlaProProArgPheIysProPheSerCysLeuSerLeuProSerSerTr 244
|||
Qy 437 GGACTAAGCGCGCTGCACACACCCCGCTAAATTTTGTATTTTAAATAAGACGGG 496
|||
Db 244 pAspTyrArgArgProProHisProAlaAsnPhePheValPheLeuValGluThrG 264
|||
Qy 497 GTTT 500
|||


```
PA (HYSE-) HYSEQ INC.
XX
PI Tang YT, Liu C, Drmanac RT;
XX
XX WPI; 2001-457603/49.
XX N-PSDB; AAH99337.
XX
XX Isolated human polynucleotides encoding polypeptides, useful for the
XX treatment and diagnosis of e.g. cancer, ulcers and HIV infection.
XX
XX Claim 20; Page 198; 1217pp; English.
XX
XX AAH99166 to AAH99904 encode the human proteins given in AAH25225 to
XX AAH25963. The proteins can have activities based on the tissues and cells
XX they are expressed in, such as: anti-inflammatory; antirheumatic;
XX antiallergic; immunosuppressive; antibacterial; endocrine; cardiac;
XX central nervous system; virucide; anti-HIV; fungicide; antitumagen;
XX cardiovascular; antianaemic; antiaggregant; haemostatic; vulnery;
XX antidiabetic; osteopathic; dermatological; antiallergic; antiasthmatic;
XX antiparkinsonian; and immunostimulant. The proteins and polynucleotides
XX encoding them can be used in gene therapy, antisense therapy and vaccine
XX production. The proteins and polynucleotides are useful for screening for
XX agonists or antagonists of a protein and for the treatment and diagnosis
XX of disorders associated with the activity of a protein e.g. inflammation,
XX rheumatoid arthritis, septic shock, pancreatitis, cardiac dysfunction,
XX neuropathology, cardiac anaphylaxis, viral, bacterial, HIV and fungal
XX infections, autoimmunity, genetic diseases, haematopoietic disorders,
XX anaemia, platelet disorders, thrombocytopaenia, wounds, burns, ulcers,
XX osteoporosis, severe combined immunodeficiency, eczema, allergic
XX rhinitis, asthma, diabetes, cancer, multiple sclerosis, depression,
XX Alzheimer's disease, Parkinson's disease, neurodegenerative and
XX neurological disorders
XX
XX SQ Sequence 213 AA;

Alignment Scores:
Pred. No.: 1.18e-35 Length: 213
Score: 378.00 Matches: 94
Percent Similarity: 63.6% Conservative: 11
Best Local Similarity: 57.0% Mismatches: 41
Query Match: 42.0% Indels: 20
DB: 4 Gaps: 4

US-10-607-806-1_COPY_5000_5500 (1-501) x AAH25396 (1-213)

QY 12 TCTGTCGCCAGGTGGAGTGGGTGGTGCATCATCATCATCATCATCATCTCTCT 71
Db 37 SerValAlaGlnAlaGluValGlnTrpHisAspLeuArgSerLeuGlnAlaProProPro 56
QY 72 GGGTTCAAAATAATCTCTTGCTCAGCCTCTCAAGTAGCTGGGACTTGTAGCTCAAGT 131
Db 57 GlyPheThrProPheSerCysLeuSerLeuProGlySerTrpAspTyrArgCys----- 74
QY 132 AGCTGGCACACACACACACACACACACACACACACACACACACACACACACACAG 190
Db 75 -----ProProProArgProAlaAsnPheLeu-----TyrPhe*****ArgArg 89
QY 191 GTTTTCACCATGTTGGCAGCTGGTGCACCTC-----CCTTTAGATTCCTCATCC 244
Db 90 GlyPheThrValLeuAlaArgMetValSerIleSer***Pro-ArgAspProProAlaSe 109
QY 245 TGCTCTATTCTCCCTTTCTAATGACGATATCCAGTTCTTCCTTACTTATCATATTATTAT 304
Db 109 rAla-----SerGlnSerAlaGlyIleThrValLeuSe 120
QY 305 TATTCTTATTATTATGACACAGAGTCTTGTCTTGTCCCAAGGCTGGAGTACAGTGGTG 364
Db 120 rLeuPhePhePhePheGluMetGluSerCysSerValAlaGlnAlaGlyValGlnTrpAr 140
QY 365 CGATCTCGGCTACATGCAAGCTCCACCTGCTGGGTTCACGCCATTCCTCCGCCCTCAGCCT 424
Db 140 gTy-rLeuGlySerLeuGlnAlaLeuProProGlyPheThrProPheSerCysLeuSerLe 160

QY 425 CCCAGTAGCTGGGACTAAAGCGCGCTGCCACACGCGCCCGCTAATTTTGTATTTT 484
Db 160 uProSerSerTrpAspTyrArgArgProProProArgProAlaAsnPhePheValPhe 180
QY 485 AATAAAGACGGG 497
Db 180 uValGluThrGly 184

RESULT 6
ADS11829
ID ADS11829 standard; protein; 186 AA.
XX
XX ADS11829;
XX
XX 16-DEC-2004 (first entry)
XX
XX Human therapeutic contig protein - SEQ ID 2066.
XX
XX antiinflammatory; neuroprotective; antianaemic; cytostatic; vulnery;
XX inflammatory; haematopoiesis; immunity; neurodegenerative; stem cell;
XX aplastic anaemia; cancer; wound healing; gene therapy.
XX
XX Homo sapiens.
XX
XX Key Location/Qualifiers
XX Misc-difference 1. 186
XX /label= Unknown, OTHER
XX /note= "OTHER = in-frame STOP codon"
XX
XX WO2004080148-A2.
XX
XX 23-SEP-2004.
XX
XX 30-SEP-2003; 2003WO-US030720.
XX
XX 02-OCT-2002; 2002US-0416186P.
XX (NUVE-) NUVELO INC.
XX
XX Tang YT, Asundi V, Ren F, Zhang J, Wehrman T, Wang Z, Ma Y;
XX Wang D, Chen R, Zhao QA, Wang J, Ghosh M, Xue AJ, Weng G, Zhou P;
XX
XX WPI; 2004-668857/65.
XX N-PSDB; ADS11231.
XX
XX New polynucleotide, useful in preparing a composition for diagnosing or
XX treating inflammatory, neurodegenerative or stem cell disorders, e.g.,
XX aplastic anemia or cancer for promoting wound healing.
XX
XX Example 2; SEQ ID NO 2066; 718pp; English.
XX
XX The invention relates to a novel isolated polynucleotide and the encoded
XX polypeptide. The molecules of the invention demonstrate antiinflammatory,
XX neuroprotective, antianaemic, cytostatic and vulnery activities and may
XX be useful in preparing a composition for diagnosing or treating
XX inflammatory, haematopoietic, immune, neurodegenerative or stem cell
XX disorders, such as aplastic anaemia or cancer, as well as for promoting
XX wound healing. The molecules may also be utilised during gene therapy
XX procedures. The current sequence is that of a human therapeutic contig
XX protein of the invention.
XX
XX SQ Sequence 186 AA;

Alignment Scores:
Pred. No.: 3.91e-35 Length: 186
Score: 373.50 Matches: 93
Percent Similarity: 60.5% Conservative: 11
Best Local Similarity: 54.1% Mismatches: 51
Query Match: 41.5% Indels: 17
DB: 8 Gaps: 4

US-10-607-806-1_COPY_5000_5500 (1-501) x ADS11829 (1-186)
```


XX Novel human secreted protein #70.
 DE Human; vaccination; gene therapy; nutritional supplement;
 KW stem cell proliferation; haematopoiesis; nerve tissue regeneration;
 KW immune suppression; immune stimulation; anti-inflammatory; leukaemia.
 OS Homo sapiens.
 PN WO200179449-A2.
 XX 25-OCT-2001.
 PF 16-APR-2001; 2001WO-US008656.
 XX 18-APR-2000; 2000US-00552929.
 PR 26-JAN-2001; 2001US-00770160.
 XX (HYSE-) HYSEQ INC.
 PA Tang YT, Liu C, Drmanac RT;
 PI WPI; 2001-611725/70.
 DR Nucleic acids encoding a range of human polypeptides, useful in genetic
 PT vaccination, testing and therapy.
 XX Claim 20; Page 168; 765pp; English.
 XX The invention relates to novel human secreted polypeptides. The
 CC polypeptides and antibodies to the polypeptides are useful for
 CC determining the presence of or predisposition to a disease associated
 CC with altered levels of polypeptide. The polypeptides are also useful for
 CC identifying agents (agonists and antagonists) that bind to them. Cells
 CC expressing the proteins are useful for identifying a therapeutic agent
 CC for use in treatment of a pathology related to aberrant expression or
 CC physiological interactions of the polypeptide. Vectors comprising the
 CC nucleic acids encoding the polypeptides and cells genetically engineered
 CC to express them are also useful for producing the proteins. The proteins
 CC are useful in genetic vaccination, testing and therapy, and can be used
 CC as nutritional supplements. They may be used to increase stem cell
 CC proliferation, to regulate haematopoiesis; and in bone, cartilage, tendon
 CC and/or nerve tissue growth or regeneration; immune suppression and/or
 CC stimulation; as anti-inflammatory agents; and in treatment of leukaemias.
 CC AAU29510-AAU3304 represent the amino acid sequences of novel human
 CC secreted proteins of the invention
 XX Sequence 324 AA;
 SQ
 Alignment Scores:
 Pred. No.: 7,99e-35 Length: 324
 Score: 371.50 Matches: 95
 Percent Similarity: 60.4% Conservative: 15
 Best Local Similarity: 52.2% Mismatches: 44
 Query Match: 41.3% Indels: 28
 DB: 4 Gaps: 5
 US-10-607-806-1_COPY_5000_5500 (1-501) x AAU29579 (1-324)
 QY 15 GTCGCCAGGCTGGAGTGGGCTGATCTCAGATCACTACAA---CCTCCATCTCT 71
 Db 140 ValAlaGlnAlaGlyValGlnTrpArgAspLeuGlySerLeuGlnProProPro 159
 QY 72 GGGTTCAATAATCTCTTGGCTCAGCCTCTCAGTACCTGGACTTGTCTCAAGT 131
 Db 160 ArgPheLysGlnPheSerCysLeuSerLeuProSerThrTrp----- 173
 QY 132 AGCTGG-----CACACACACCATGCCAGCTAATTTTGTGTG----- 171
 Db 174 SerTrpValTrpArgHisAlaProCysProAlaAsnPheVal-PhePhePhePhe 193
 QY 172 -----TTTTTTTGGTAGACAGAGGTTTTCACATGTTGGCAGGCTG----- 214
 ||||||||||||||||| ||| ||||||||| |||||

Db 193 ePhePhePheLeuValGluLysGly-PheSerMetLeuLeuArgLeuValLeuAsnSer* 213
 QY 215 -----GGTGACCTCCCTTTTAGATTTCTCTCATCTCTGCTCTATTCTTCCCTTT 263
 Db 213 **ProHisAspProAspProAlaSerAlaSerGlnSerAlaGlyllehrGlyValS 233
 QY 264 CTAATGCAATATCCAGTTCTTCTTATCATCATTTATTATTATTATTATTATTGAG 323
 Db 233 erHisHisThrArgProMetSerPheLysAsnIleTrpPhePhePhePhePhe-Glu 252
 QY 324 ACAGAGTCTTCTTGTGCGCAAGCTGAGTACAGTGTGGGATCTCGGCTCATCGAA 383
 Db 253 ThrGluSerArgSerValAlaGlnAlaGlyValGlnTrpArgAspLeuSerArgGln 272
 QY 384 GCTCCACCTGCTGGGTTTCCAGCCATTCTCCGCTCAGCCTCCCGAGTAGCTGGGACTAA 443
 Db 273 ProProProGlyPheLysArgPheSerCysLeuSerLeuSerSerTrpAspTrp 292
 QY 444 AGGCGCTGCCACACCGCCGCTAATTTTGTATTATTTTAAATAAGACGGGTT 499
 Db 293 ArgArgValProProMetProGly-***PheCysAllePheArgArgAspGlyVal 310
 RESULT 9
 ABB11464
 ID ABB11464 standard; peptide; 241 AA.
 XX
 AC ABB11464;
 XX
 DT 11-JAN-2002 (first entry)
 XX Human neuronal thread protein homologue, SEQ ID NO:1834.
 DE Human; cytokine; cell proliferation; cell differentiation; growth factor;
 KW haematopoiesis regulation; tissue growth; immunomodulator; activin;
 KW inhibin; chemotaxis; chemokinesis; thrombolysis; oncogenesis;
 KW proliferation; metastasis; cancer; tumour; haematopoietic disorder;
 KW myeloid cell disorder; lymphoid cell disorder; asthma; arthritis;
 KW chronic inflammatory condition; proliferative retinopathy;
 KW atherosclerosis; coronary heart disease; arterial ischaemia;
 KW bone disorder; osteoporosis; vascular growth disorder; immune disorder;
 KW tissue regeneration; wound healing; infection; immune disorder;
 KW cell culture; drug screening; gene therapy; anti-inflammatory;
 KW antiasthmatic; antiarthritic; haemostatic; antiarteriosclerotic;
 KW cytosatic; osteopathic; vasotropic; cardiant; virucide; antibacterial;
 KW antifungal; vulnery; antiulcer.
 XX Homo sapiens.
 OS
 XX WO200157188-A2.
 FN
 PD 09-AUG-2001.
 XX
 PF 05-FEB-2001; 2001WO-US003800.
 XX
 PR 03-FEB-2000; 2000US-00496914.
 PR 27-APR-2000; 2000US-00560875.
 XX
 PA (HYSE-) HYSEQ INC.
 XX
 XX Tang YT, Liu C, Drmanac RT;
 PI WPI; 2001-457740/49.
 XX N-PSDB, ABA08708.
 DR Human proteins and DNA encoding sequences useful for preventing, treating
 PT or ameliorating a medical condition in a mammalian subject e.g. arthritis
 PT and cancer.
 XX Claim 20; Page 194; 1963pp; English.
 PS
 XX Sequences ABB10981-ABB12330 represent 1350 novel human polypeptides, and
 CC sequences ABA08225-ABA09574 represent nucleic acids encoding them. The
 CC invention also relates to vectors and recombinant host cells comprising a

CC nucleotide of the invention, methods of producing the novel polypeptides,
 CC antibodies against the polypeptides, methods of detecting the nucleotides
 CC or polypeptides in a sample, and methods of identifying compounds which
 CC bind to polypeptides of the invention. Although novel, many of the
 CC polypeptides of the invention have homology to known proteins, thereby
 CC giving an insight into their probable biological activities, and hence
 CC potential therapeutic applications. The polypeptides of the invention may
 CC have various activities, including cytokine, cell proliferation or cell
 CC differentiation activities; stem cell growth factor activity;
 CC haematopoietic regulatory activity; tissue growth activity;
 CC immunomodulatory activity; activin- or inhibin-related activities;
 CC chemotactic or chemokinetic activities; haemostatic, thrombotic or
 CC thrombolytic activities; receptor or ligand activities; or may be
 CC involved in oncogenesis, cancer cell proliferation or metastasis.
 CC Depending on their biological activities, polypeptides and nucleotides of
 CC the invention are useful for preventing, treating or ameliorating medical
 CC conditions, e.g., by protein or gene therapy. Such conditions include
 CC cancers, haematopoietic disorders (e.g., myeloid or lymphoid cell
 CC disorders), chronic inflammatory conditions (e.g., asthma or arthritis),
 CC proliferative retinopathy, atherosclerosis, coronary heart disease,
 CC arterial ischaemia, bone disorders (e.g., osteoporosis), and abnormal
 CC vascular growth. Polypeptides involved with tissue regeneration and
 CC repair (or nucleic acids encoding them) may be used to promote wound
 CC healing (e.g., of burns, incisions and ulcers), while those with
 CC immunomodulatory activities may be used in the treatment of viral,
 CC bacterial and fungal infections in addition to immune disorders.
 CC Polypeptides with growth factor activity may be used in cell cultures to
 CC promote cell growth. For example, such polypeptides may be used to
 CC manipulate stem cells in culture to give rise to neuroepithelial cells
 CC that can be used to augment or replace cells damaged by illnesses.
 CC autoimmune disease or accidental damage. The polypeptides and nucleotides
 CC may also be used in the diagnosis of the above conditions, and in drug
 CC screening techniques. The present sequence represents a novel human
 CC polypeptide of the invention
 XX
 SQ Sequence 241 AA;

Alignment Scores:
 Pred. No.: 2,89e-34 Length: 241
 Score: 366.50 Matches: 96
 Percent Similarity: 52.1% Conservative: 17
 Best Local Similarity: 44.2% Mismatches: 44
 Query Match: 40.7% Indels: 60
 DB: 4 Gaps: 4

US-10-607-806-1_COPY_5000_5500 (1-501) x ABB11464 (1-241)

QY 6 GCTCACTCTGCGCCAGCTGGAGTGGCGTGGATCTCAGATCACTACACCTCCA 65
 Db 8 SerArgSerValAlaGlnAlaGlyValGlnTrpCysAsnLeuGlySerLeuGlnAlaLeu 27
 QY 66 TCTCCTGGGTTCAATAATCTCTGCTCAGCTCTCAAGTAGCTGGGACTTGTAGCTC 125
 Db 28 ProProGlyPhe-SerHisSerProAlaSerAlaSerArgValAlaGlyThr----- 44
 QY 126 TCAAGTAGCTGGCACACACACACAGCCAGCTAATTTTGTGTGTTTTTGTGTAGA 185
 Db 45 -----ThrGlyThrArgHis***AlaArgLeuIlePhe--TyrIle---PheSerArg 60
 QY 186 GACAGGTTTTACCATGTTGGCCAGGCTGGGTGACCTCC----- 224
 Db 61 AspGlyValSerProCys***ProGlyTrpSer***SerProAspLeuValIleArgPro 80
 QY 225 -----CTTTAGATTCTCCTCATCTGCTCTATTCTTC 257
 Db 81 ProArgLeuProLysCysTyrAspTyrArgArgGluProProArgProAla***PhePhe 100
 QY 258 CCCTTTCTAATGAGTATCCAGTTTCCTTACTT----- 290
 Db 101 ValPheLeuValGluGlnGlyPheThrMetLeuAlaArgMetValSerIleSer***Pro 120
 QY 290 ----- 290

Db 121 GlnCysAspLeuProAlaSerValSerGlnAsnAlaGlyIleThrGlyValSerHisCys 140
 QY 291 -----ATCACATTATTATTATTCTTATTATTATTGAGACAGAGCTCTGCTTT 338
 Db 141 AlaTrpProCysLeuHisPheCysPheGlyPhePheGlyPhePheGluMetGluSerCysSer 160
 QY 339 GTCCCAAGGCTGGAGTACAGTGTGGATCTCGGCTCACTGCAAGCTCCACCTGCTGGG 398
 Db 161 ValAlaGlnAlaGluValGlnTrpHisAspLeuArgSerLeuGlnAlaProProGly 180
 QY 399 TTCAGCGCATTTCTCCGCCCTCAGCTCCCAAGTAGCTGGGACTAAAGCGCTGCCACCA 458
 Db 181 PheThrProPheSerCysLeuSerLeuProGlySerTrpAspTyrArgArgProProPro 200
 QY 459 CGCCCGCTAATTTTGTATTTTAAATAAGACGGGTT 499
 Db 201 ArgProAlaAsn--PheCysIlePheSerArgAspGlyVal 213

RESULT 10

ABG68738
 ID ABG68738 standard; protein; 361 AA.

AC ABG68738;

DT 07-OCT-2002 (first entry)

XX Human prostate specific protein DEX0293_107.

XX Human; prostate specific nucleic acid; PSNA; prostate cancer; PSP;
 KW prostate specific protein; cytostatic; non-cancerous prostate disease;
 KW gene therapy; cancer; immunostimulant; vaccine.

XX Homo sapiens.

XX WO200255735-A2.

XX 18-JUL-2002.

XX 27-NOV-2001; 2001WO-US044363.

XX 27-NOV-2000; 2000US-0253176P.

PA (DIAD-) DIADEXUS INC.

PI Salceda S, Macina RA, Recipon H, Cafferkey R, Ali S, Sun Y;
 PI Liu C, Chen S;

DR WP1; 2002-557831/59.

DR N-PSDB; ABK97631.

XX New prostate specific genes, useful for treating or diagnosing cancer, or
 PT useful as vaccines for treating cancer, particularly prostate cancer, in
 PT a patient.

XX Claim 11; Page 207-208; 212pp; English.

XX The invention relates to a new isolated prostate-specific nucleic acid
 CC (PSNA) molecule comprising the cDNA sequences appearing as ABK97574-
 CC ABK97642 which encode prostate specific proteins appearing as ABG68701-
 CC ABG68746, or a sequence hybridising to a PSNA or which has 60% sequence
 CC homology with a PSNA. Also included are a method of determining the
 CC presence of a PSNA in a sample, a vector comprising the PSNA, a host cell
 CC comprising the vector, producing the polypeptide encoded by the PSNA, a
 CC method of determining the presence of a PSP in a sample, diagnosing and
 CC monitoring the presence and metastases of prostate cancer in a patient,
 CC kit for detecting a risk of cancer or presence of cancer in the PSNA or
 CC (the kit comprising a means for determining the presence of the PSNA or
 CC PSP in a sample of a patient) and a vaccine comprising the polypeptide or
 CC the nucleic acid encoding the polypeptide. The PSNA, PSP and anti-PSP
 CC antibody are useful for diagnosing and treating cancer in a patient (e.g.
 CC by gene therapy). The nucleic acid molecule and polypeptide are also
 CC useful as vaccines for treating cancer, particularly prostate cancer and
 CC non-cancerous prostate diseases. The present sequence is a PSP of the

```

CC invention
SQ Sequence 361 AA;

Alignment Scores:
Pred. No.: 3-26e-34 Length: 361
Score: 366.50 Matches: 96
Percent Similarity: 60.4% Conservative: 17
Best Local Similarity: 51.3% Mismatches: 44
Query Match: 40.7% Indels: 30
DB: 5 Gaps: 6

US-10-607-806-1_COPY_5000_5500 (1-501) x ABG68738 (1-361)
QY 9 CACTGTGCGCCAGCTGGAGTGGCGGTGATCTCAGATCACTCAACCTCCATCT 68
   |||||
DB 171 HisSerValAlaGlnAlaGlyValGlnTrpCysAenLeuGlySerLeuGlnProLeuPro 190
QY 69 CTGGGTTCAATAATTTCTTGTCTCAGCTCTCAAGTAGCTGGAGCTTAGCTCTCA 128
   |||||
DB 191 ProArgPheArgPheSerCysLeuSerLeuLeuSerSerTrpAsp----- 206
QY 129 AGTAGCTGCACACACACACACACGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 185
   |||||
DB 207 ---HisArgHisAlaProProCysLeuAlaAenPheLeu-PhePheLysPheLeuValAs 225
QY 186 GACAGGTTTTCACCATGTTGGCCAGCTG-----GGTGACCTCC 224
   : |||
DB 225 p-GlnSerPheThrMetLeuAlaAArgLeuValLeuAenSerAlaProSerGlyAspLeuP 245
QY 225 CTTTGTAGATTCCTCATCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 278
   |||||
DB 245 roAlaProAlaSerGlnSerAlaGlyIleThrGlyValArgHisCysThrTrpProLysS 265
QY 279 GTTTC-----CTTACTTATCAC-----ATTATATTATT 308
   |||||
DB 265 erPheLysPheAlaAspSerHisIleGlyLeuAlaPheHisPheAlaPhePhePheP 285
QY 309 CTTATTATTATGACAGAGCTCTGTTGTCGCCAAGGCTGGAGTACAGTGGTGCAT 368
   : |||||
DB 285 hePhePhePhe-AlaValAlaSerHisProIleAlaGlnAlaGlyValGlnTrpArgAsp 304
QY 369 CTCGGCTCAGTCGACAGCTCCACCTGCTGGTTCAGCCCATCTCCCGCTCAGCCTCCC 428
   |||||
DB 305 LeuGlySerLeuGlnProProProGlyPheLysGlnPheLeuCysLeuSerLeuPro 324
QY 429 AGTAGCTGGACTAAGCGGCTGCGCACACGCGCGCTAATTTTGTATTATTATA 488
   |||||
DB 325 GlySerTrpAspTrpArgAlaProProArgGlnAlaAen--PheCysIlePheSerA 344
QY 489 AAGACGGGGTT 499
DB 344 rGAspGlyVal 347

RESULT 11
AAU30235
ID AAU30235 standard; protein; 381 AA.
XX
AC AAU30235;
XX
DT 18-DEC-2001 (first entry)
XX
DE Novel human secreted protein #726.
XX
KW Human; vaccination; gene therapy; nutritional supplement;
KW stem cell proliferation; haematopoiesis; nerve tissue regeneration;
KW immune suppression; immune stimulation; anti-inflammatory; leukaemia.
XX
OS Homo sapiens.
XX
FN WO200179449-A2.
XX
PD 25-OCT-2001.

```

```

XX
PF 16-APR-2001; 2001WO-US008656.
XX
PR 18-APR-2000; 2000US-00552929.
XX
PR 26-JAN-2001; 2001US-00770160.
XX
PA (HYSE-) HYSEQ INC.
XX
PI Tang YT, Liu C, Drmanac RT;
XX
DR WPI; 2001-611725/70.
XX
PT Nucleic acids encoding a range of human polypeptides, useful in genetic
PT vaccination, testing and therapy.
XX
PS Claim 20; Page 264-265; 765pp; English.
XX
CC The invention relates to novel human secreted polypeptides. The
CC polypeptides and antibodies to the polypeptides are useful for
CC determining the presence of or predisposition to a disease associated
CC with altered levels of polypeptide. The polypeptides are also useful for
CC identifying agents (agonists and antagonists) that bind to them. Cells
CC expressing the proteins are useful for identifying a therapeutic agent
CC for use in treatment of a pathology related to aberrant expression or
CC physiological interactions of the polypeptide. Vectors comprising the
CC nucleic acids encoding the polypeptides and cells genetically engineered
CC to express them are also useful for producing the proteins. The proteins
CC are useful in genetic vaccination, testing and therapy, and can be used
CC as nutritional supplements. They may be used to increase stem cell
CC proliferation; to regulate haematopoiesis; and in bone, cartilage, tendon
CC and/or nerve tissue growth or regeneration; immune suppression and/or
CC stimulation; as anti-inflammatory agents; and in treatment of leukaemias.
CC AAU29510-AAU33304 represent the amino acid sequences of novel human
CC secreted proteins of the invention
XX
SQ Sequence 381 AA;

Alignment Scores:
Pred. No.: 3.8e-34 Length: 381
Score: 366.00 Matches: 89
Percent Similarity: 56.0% Conservative: 13
Best Local Similarity: 48.9% Mismatches: 43
Query Match: 40.7% Indels: 37
DB: 4 Gaps: 7

US-10-607-806-1_COPY_5000_5500 (1-501) x AAU30235 (1-381)
QY 8 TCACCTGTGCGCCAGGCTGGAGTGGCGGTGATCTCAGATCACTCAACCTCCATC 67
   |||||
DB 103 SerLeuCysTrpProGlyTrpSerArgValAla***SerArgIle---ThrAlaThr-Se 121
QY 68 TCCTGGGTTCAATAATTTCTTGCCTCAGCTCTCAAGTAGCTGGAGCTTAGCTCTC 127
   |||||
DB 121 rProGlyLeuLys***PheAlaCysPheSerLeuProSerSerArgAspTrp----- 138
QY 128 AAGTAGCTGGCACACACACACACCATGCGCGGTGAGCTCCCTTTTAGATTCTCTCATCCTGC 187
   |||||
DB 139 -----ArgHisValProProHisProGlyAsnPhe---CysIle---PheGlyArgAs 154
QY 188 CAGGTTTTCACCATGTTGGCCAGGCTGGGTGAGCTCCCTTTTAGATTCTCTCATCCTGC 247
   |||||
DB 154 pGluValSerProCysTrpProGlyTrpPhe***Thr-----ProAs 168
QY 248 TCTATTCTTCCCTTTCTAATGAGTATCCAGTATCCAGTTTCTCTTA----- 287
   |||||
DB 168 pLeuArgTrpPro-----ProAlaSerAlaSerGlnSerAlaGluIleI 183
QY 288 -----CTTATCACATTTATTATTATTATTATTATTATTATTAT 316
   |||||
DB 183 eGlyValSerHisHisThrTrpProGlnGluValPheLeuPheLeuAsnLeuPheIety 203
QY 317 TATTGACACAGAGTCTTGTCTTCGCCAAGGCTGGAGTACAGTGGTGCATCTCGGCTC 376
   |||||

```


Db 203 rLeuArgTrpSerLeuaspSerValAlaGlnAlaArgValGlnhArgaspLeuGlySe 223
Qy 377 ACTGACAGCTCCAGCTGCTGGGTTACGCCATTCTCCGCTCAGCTCCGCCAGTGTG 436
Db 223 rLeuGlnAlaProProArgPheLysPheProPheSerCysLeuSerLeuProSerSerTr 243
Qy 437 GGACTAAAGGCGCTGCCACACCGCCCGCTAAATTTTCTGTTATTTTAAATAAGACGGG 496
Db 243 pAspTyrArgArgProProHisProAlaAsnPheValPheLeuValGluThrGI 263
Qy 497 GTTT 500
Db 263 yPhe 264
RESULT 12
ID ADC87151 standard; protein; 217 AA.
XX ADC87151;
XX
DT 01-JAN-2004 (first entry)
XX
DE Human GPCR protein SEQ ID NO:1604.
XX
KW human; GPCR; guanosine triphosphate-binding protein coupled receptor;
KW gene therapy.
XX
OS Homo sapiens.
XX
PN EP1270724-A2.
XX
PD 02-JAN-2003.
XX
PF 18-JUN-2002; 2002EP-00013517.
XX
PR 18-JUN-2001; 2001JP-00246789.
XX
PA (NAAD-) NAT INST ADVANCED IND SCI & TECHNOLOGY.
PA (ADSC-) CENT ADVANCED SCI & TECHNOLOGY INCUBATIO.
PI Suwa M, Asai K, Akiyama Y, Aburatani H;
XX
DR WPI; 2003-315783/31.
DR N-PSDB; ADC87150.
XX
PT New polynucleotide, useful for preparing a composition for treating a
PT patient in need of increased or suppressed activity or expression of the
PT guanosine triphosphate-binding protein coupled receptor.
XX
PS Claim 2; SEQ ID NO 1604; 28pp; English.
XX
CC The invention relates to a novel polynucleotide encoding a guanosine
CC triphosphate-binding protein coupled receptor (GPCR). A polynucleotide of
CC the invention may have a use in gene therapy. The polynucleotide and
CC polypeptide are useful for preparing a composition for treating a patient
CC in need of increased or suppressed activity or expression of the
CC guanosine triphosphate-binding protein coupled receptor. The protein
CC sequences shown in ADC85549-ADC87617 represent GPCR's of the invention.
XX
SQ Sequence 217 AA;
Alignment Scores:
Pred. No.: 7.32e-34 Length: 217
Score: 363.00 Matches: 96
Percent Similarity: 55.2% Conservative: 5
Best Local Similarity: 52.5% Mismatches: 38
Query Match: 40.3% Indels: 44
DB: 7 Gaps: 4
US-10-607-806-1_COPY_5000_5500 (1-501) x ADC87151 (1-217)
Qy 7 CTCACCTCTGCGCCAGGCTGGAGTGGGTGGCGGTATCTCAGATCACTACACCTCCAT 66
|||||

Db 55 LeuThrLeuLeuProArgLeuGluCysSerGlyThrIleSerAlaHisCysAsnLeuHis 74
Qy 67 CTCCTGGTTCAATAATCTCTTGCCTCAGCTCTCAAGTAGCTGGGACTTGTAGCTCT 126
Db 75 LeuProGlySerGlyAspSerHisThrSerAlaSerGlnValAlaGly----- 90
Qy 127 CAAGTCTGGCACACACACACCATGCCAGCTAAATTTTGTGTGTTTTTTTGGTAGAG 186
Db 91 ---IleThrGlyAlaHisArgHisAlaTrpLeuIlePhe-----ValPheLeuLeuGlu 107
Qy 187 ACAGTTTTCACCATGTTGGCCAGGCTGGGTGAGCTCCCTTTTATGATTCT----- 236
Db 108 ThrGlyPheCysHisValGlyGlnAlaGly--LeuGluLeuLeuThrSerGlyAspProp 127
Qy 237 -----CCTCATCTGCTCTATTCTCC 258
Db 127 roAlaLeuAlaSerGlnSerAlaGlyIleThrGlyValSerHisProLysPhePhep 147
Qy 259 CCTTTCTAATGCAGTATCCAGTTTCTTACTTATCATATTTATTATTATTATTATTA 318
Db 147 hePhe-----Phep 150
Qy 319 TTGAGACAGAGTCTTGTCTTGTTCGCCAAGCTGAGTACAGTGTGGATCTCGGCTCAC 378
Db 150 heGluThrGluSerHisThrValThrTrpAlaGlyValGlnTrpCysAspLeuGlySerL 170
Qy 379 TGAAGCTCCACCTGCTGGTTTACGCCATTCTCCGCTCAGCTCCAGTACAGTACAGTGGG 438
Db 170 euGlnProProProGlySerSerAspSerProAlaSerAlaSerArgValAlaGly 189
Qy 439 ACTAAGCGCTGCCACCGCTAAATTTTGTATTTTGTATTTTAAATAAGACGGGT 498
Db 190 IleThrGlyAlaHisHisAlaGlnLeuIlePheLeuPhe-PheArgArgaspGlyVa 209
Qy 499 T 499
Db 209 I 209
RESULT 13
ID ADR09211 standard; protein; 179 AA.
XX
AC ADR09211;
XX
DT 04-NOV-2004 (first entry)
XX
DE Human protein useful for treating neurological disease Seq 2717.
KW human; oligo-capping method; diagnostic marker; gene therapy;
KW osteoporosis; neurological disease; Alzheimer's disease;
KW Parkinson's disease; dementia; short memory; cancer;
KW sense or motor function; emotional reaction; fear response; panic;
KW osteopathic; neuroprotective; nootropic; antiparkinsonian; cyostatic;
KW tranquiliser.
XX
OS Homo sapiens.
XX
PN EP1447413-A2.
XX
PD 18-AUG-2004.
XX
PF 12-FEB-2004; 2004EP-00003145.
XX
PR 14-FEB-2003; 2003JP-00102207.
XX
PR 09-MAY-2003; 2003JP-00131452.
XX
PA (REAS-) RES ASSOC BIOTECHNOLOGY.
XX
PI Isogai T, Yamamoto J, Nishikawa T, Isono Y, Sugiyama T, Otsuki T;
PI Wakamatsu A, Ishii S, Nagai K, Irie R;
XX
DR WPI; 2004-583265/57.
DR N-PSDB; ADR07255.

XX New 1995 cDNA, useful for treating osteoporosis, neurological diseases,
PT Alzheimer's diseases, Parkinson's diseases, dementia and various cancers.
PS
XX
XX Claim 1; SEQ ID NO 2717; 2686pp; English.
XX
XX This invention relates to novel, isolated full length human cDNA
CC molecules and the encoded proteins thereof. Specifically, it refers to
CC cDNA clones obtained by an oligo-capping method, where none of these
CC clones are identical to any known human mRNAs. The present invention
CC describes an immunoassay to identify agonists and antagonists, as well as
CC antibodies, antisense molecules and siRNAs that can all be used to bind
CC to and modulate expression of the cDNA molecules. As such, these
CC molecules are useful for diagnostic markers or therapeutic targets for
CC the various diseases or morbid states. In particular, they are useful in
CC gene therapy for treating osteoporosis, neurological disease, Alzheimer's
CC disease, Parkinson's disease, dementia, short memory and various cancers,
CC as well as for maintaining equilibrium of sense or motor function, and
CC for treating emotional reaction, fear response and panic. Accordingly,
CC they exhibit osteopathic, neuroprotective, nootropic, antiparkinsonian,
CC cytosolic and tranquiliser activities. This polypeptide is a protein
CC encoded by a full length human cDNA sequence of the invention. NOTE: This
CC sequence is not given in the sequence listing of the specification but
CC can be obtained on CD-ROM from the European Patent Office, Vienna Sub-
CC office.

SQ Sequence 179 AA;

Alignment Scores:
Pred. No.: 2,398-33 Length: 179
Score: 358.50 Matches: 92
Percent Similarity: 61.3% Conservative: 11
Best Local Similarity: 54.8% Mismatches: 43
Query Match: 39.8% Indels: 22
DB: 8 Gaps: 5

US-10-607-806-1_COPY_5000_5500 (1-501) x ADR09211 (1-179)

Qy 12 TCTGTGCGCCAGGTGGAGTGGGTGGGTGATCTCAGATCACTACACCTCCATCTCCT 71
Db 5 SerValSerGlnAlaGlyValGlnTrpLeuAspLeuSerSerLeuGlnProProSer 24
Qy 72 GGGTTCAATAATTTCTTGGCTCAGCTCTCAAGTAGCTGGAGCTGTAGCTCTCAAGT 131
Db 25 GlyPheLysArgPheSerCysLeuSerLeuGlnSerSerTrpAspTyr----- 40
Qy 132 AGCTGGCACACACACCATGCCCCAGCTAATTTTGTGTGTTTTTGTGTAGACAGG 191
Db 41 ---ArgHisAlaProProCysProAlaGly-PheCysLeu-----CysValGluThrGl 57
Qy 192 TTTTCACCATGTTGGCCAGGTGGGTGGACCTCCCTTTTAGATTCTCTCCTCCTGCTCTA 251
Db 57 yPheHisHisValGlyGlnAlaSer-----LeuLysLeuLeuThrSe 71
Qy 252 TCTTCCCTTCTTAATGCAATGCCAGTTCCATTTCTTACTATPCA-----CATTTATT 302
Db 71 rSerAspProProAlaLeuAlaSerGlnIleAlaGlyIleThrGlyMetIleHisAl 91
Qy 303 ATATTCTTATTATTATGACAGAGCTGTGCTTTGTGCCAAGGTGGAGTACAGTGG 362
Db 91 aGlnProLeuLeuPheLeuArgGlnGlyValThrLeuLeuProArgLeuGluCysSerSe 111
Qy 363 TGGCATCTCGGCTCACTGCAAGCTCCACCTCGTGGTTTCAGCCCATTTCTCCGCTCAGC 422
Db 111 rAlaIleLeuAlaHisCysIleLeuCysLeuProGlySerSerAspSerProAlaSerAl 131
Qy 423 CTCCCCAGTAGCTGGGACTAAGCGGCTGTGCACACGCCCGCTAATTTTGTGATTT 482
Db 131 aSerGlnValAlaGlyThrThrGlyValCys-----ProleuIle--PheValPhe 147
Qy 483 TTAATAAGACGGGGTTT 500
Db 148 ValValGluMetGlyPhe 153

RESULT 14
ADE09985

ID ADE09985 standard; protein; 196 AA.

XX ADE09985;

XX 29-JAN-2004 (first entry)

XX Novel protein-related contig polypeptide sequence #573.

XX novel gene; novel protein; tissue marker; molecular weight marker;
KW chromosome marker; genetic disorder; contig.

XX Unidentified.

XX WO2003054152-A2.

XX 03-JUL-2003.

XX 10-DEC-2002; 2002WO-US039555.

XX 10-DEC-2001; 2001US-0339739P.

XX 11-DEC-2001; 2001US-0339453P.

XX 14-MAR-2002; 2002US-0365091P.

XX 14-MAR-2002; 2002US-0365384P.

XX 12-APR-2002; 2002US-0372381P.

XX 12-APR-2002; 2002US-0372615P.

XX 22-APR-2002; 2002US-00128558.

XX 24-APR-2002; 2002US-0376045P.

XX (HYSE-) HYSEQ INC.

XX Tang YT, Auundi V, Goodrich RW, Ren P, Zhang J, Zhao QA, Wang J;
PI Ghosh M, Xue AJ, Wehrman T, Meng G, Zhou P, Drmanac RT, Wang Z;
PI Ma Y, Wang D, Chen R, Xu C, Boyle BJ;
XX WPI; 2003-569235/53.

XX New polynucleotides, useful for expressing recombinant proteins for
PT analysis, characterization or therapeutic use, or as markers for tissues
PT in which the corresponding protein is preferentially expressed.

XX Disclosure; SEQ ID NO 3051; 1177pp; English.

XX The invention comprises the amino acid and coding sequences of novel
CC proteins. The DNA and protein sequences of the invention are useful as:
CC markers for tissues in which the corresponding protein is preferentially
CC expressed; as molecular weight markers on gels; as chromosome markers or
CC tags; to identify chromosomes or to map related gene positions; and to
CC compare with endogenous DNA sequences in patients to identify potential
CC genetic disorders. The present amino acid sequence was used in the
CC exemplification of the invention.

XX SQ Sequence 196 AA;

Alignment Scores:

Pred. No.: 6,398-33 Length: 196
Score: 355.00 Matches: 93
Percent Similarity: 62.4% Conservative: 13
Best Local Similarity: 54.7% Mismatches: 47
Query Match: 39.4% Indels: 19
DB: 7 Gaps: 4

US-10-607-806-1_COPY_5000_5500 (1-501) x ADE09985 (1-196)

Qy 12 TCTGTGCGCCAGGTGGAGTGGGTGGGTGATCTCAGATCACTACACCTCCATCTCCT 71

Db 10 SerValAlaGlnAlaGlyValGlnTrpCysAspLeuGlySerLeuGlnAlaProProPro 29

Qy 72 GGGTTCAATAATTTCTTGGCTCAGCTCTCAAGTAGCTGGAGCTGTAGCTCTCAAGT 131

Db 30 GlyPheThrProPheSerCysLeuSerLeuProSerSerTrpAspTyr----- 45

CC nucleic acids encoding the polypeptides and cells genetically engineered
CC to express them are also useful for producing the proteins. The proteins
CC are useful in genetic vaccination, testing and therapy, and can be used
CC as nutritional supplements. They may be used to increase stem cell
CC proliferation; to regulate haematopoiesis; and in bone, cartilage, tendon
CC and/or nerve tissue growth or regeneration; immune suppression and/or
CC stimulation; as anti-inflammatory agents; and in treatment of leukaemias.
CC AAU29510-AAU3304 represent the amino acid sequences of novel human
CC secreted proteins of the invention

Sequence 296 AA;

Alignment Scores:		
Pred. No.:	1.25e-32	296
Score:	353.00	98
Percent Similarity:	47.8%	Conservative: 13
Best Local Similarity:	42.2%	Mismatches: 75
Query Match:	39.2%	Indels: 5
DB:	4	Gaps: 5

UUS-10-607-806-1 COPY 5000 5500 (1-501) x AAU33304 (1-296)

QY	13	CTGTCGCCAGCGTGGAGTGGCGGTGGGTGATCTCAGATCACTACAACTCCATCTCCTCGT	72
DB	56	LeuSerProArgLeuGluCysAsnGlyValIleSerAlaHisCysAsnLeuHisProPro	75
QY	73	GGTTCAATAATTCCTTGCTCAGCTCAGCCCTCAAGTAGCTGGGACTTGTAGCTCTCAAGTA	132
DB	76	GlySerAspSerProAlaSerAlaAlaArgValAlaGly	90
QY	133	GCTGCGCACACACCACTGCCAGCTCAATTTTTTGTGTGTTT	189
DB	91	ThrGlyThrCysHisHisAlaGlnLeuIlePhePhePheValPheLeuValGluThr	110
QY	190	GGTTTTTCACCATGTGGCCAGCGTGGG	216
DB	111	GlyPheHisHisAlaAlaGlnAlaGlySerGlnThrProAspLeuArg	130
QY	217	-----TGACCTCCCTTTTAGATT	234
DB	131	LeuGlyPheProLysCys**AspTyrArgArgAlaAlaIleValProGlyIlePheLeu	150
QY	235	CTC	237
DB	151	LeuHis**IleArg**LysValProThrLeuLeuThrAspMetArgAsnAlaSerGlu	170
QY	237	-----	237
DB	171	TyrAspCysAspPheSerThrAsnLysIleAspLysGluGluThrPheSer**AsnAla	190
QY	238	-----CTCATCTGCTCTATTCTTCCCTTTCTTAATGCAGTATCAGATTCTCT---	285
DB	191	SerLeuAsnLeuCysLeuLeuSerThr**ProTyrGluMetValThrHisPheLysGly	210
QY	286	TACTTATCACATTTATTATTCTTATTATTATTATTGAGACAGAGTCTTGCTTT	344
DB	211	Tyr**IleLeuProLeuPhePhePhePheLeuArgGlnSerLeuSerValThr	230
QY	345	AAGGCTGGAGTACAGTGGTGCATCTCGGCTCACTGCAAGCTCCACTCTCTGGTTTCACG	404
DB	231	GlnAlaGlyValGlnTrpHisAsnLeuGlySerLeuGlnProLeuProProGlyPheLys	250
QY	405	CCATCTCCCGCTCAGCTCCCGAGTACTGGGACTAAAGGCGCTGCCACCAACGCCCC	464
DB	251	GlnPheSerCysLeuSerLeuProSerSerTrpAspTyrArgTyrGlnLeuProArgLeu	270
QY	465	GCTAATTTTTTGTATTTTATAAGACGGGGTTT	500
DB	271	AlaAsnPhePheValPheLeuValGluThrGlyPhe	282

Search completed: February 8, 2006, 09:59:47
Job time : 86.692 secs

[illegible]

RESULT 15

AAU33304
ID AAU33304 standard: protein: 296 AA.

XX AC AAU33304:

DT 18-DEC-2001 (first entry)

XX
DE Novel human secreted protein #3795.

Human: vaccination: gene therapy: r

KW stem cell proliferation; haematopoiesis

XX 00000000-00

2007-05-25

XX
DE
16-APR-2001: 2001WO-IIS008656XX
25 10 1000 2000 2000000-00552929

PR 26-JAN-2001; 2001US-00770160.
yy

PA (HYSE-) HYSEQ INC.
yy

PI Tang YI, Liu C, Drmanac RI;
yy

DR YY WPI; 2001-611/25/70.

PT
BT
Nucleic acids encoding a range of
vaccination testing and therapy.

XX
PS
Claim 20. Page 760: 765pp: EnglishXX The invention relates to novel hum
CC

CC potyviruses and antibodies to the
CC determining the presence of or pre-

CC identifying agents (agonists and antagonists) with altered levels of porypeptide

CC expressing the proteins are useful for use in treatment of a pathological

and, second, physiological interactions of the

THIS PAGE BLANK (USPTO)

GenCore version 5.1.1.7
Copyright (c) 1993 - 2006 Bioceleration Ltd.

OM nucleic - protein search, using frame_plus_n2p model

Run on: February 8, 2006, 09:43:51 ; Search time 3.68824 Seconds
(without alignments)
2613.961 Million cell updates/sec

Title: US-10-607-806-1_COPY_5000_5500
Perfect score: 900
Sequence: 1 acggggctcactgtcgcc.....tttaataaagacgggggttc 501

Scoring table: BLOSUM62
Xgapop 10.0 , Xgapext 0.5
Ygapop 10.0 , Ygapext 0.5
Fgapop 6.0 , Fgapext 7.0
Delop 6.0 , Delext 7.0

Searched: 283416 seqs, 96216763 residues

Total number of hits satisfying chosen parameters: 566832

Minimum DB seq length: 0
Maximum DB seq length: 2000000000
Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Command line parameters:

-MODEL=frame+ n2p.model -DEV=xlp
-Q=/abs/ABSSWEB_spool/US10607806/runat_08022006_091821_26876/app_query.fasta_1
-DB=PIR -QFMT=fastan -SUFFIX=n2p.rpr -MINMATCH=0.1 -LOOPCL=0 -LOOPEXT=0
-UNITS=bits -START=1 -END=1 -MATRIX=blosu62 -TRANS=human40.cdi -LIST=45
-DOALIGN=200 -THR SCORE=pct -THR MAX=100 -THR MIN=0 -ALIGN=15 -MODE=LOCAL
-OUTFMT=ptc -NORM=ext -HEAPSIZE=500 -MINLEN=0 -MAXLEN=200000000 -HOST=abs502p
-USER=US10607806 @CGN 1.1.92 @runat_08022006_091821_26876 -NCPU=6 -ICPU=3
-NO_WMAP -NEG SCORES=0 -WAIT -DSPBLOCK=100 -LONGLOG -DEV TIMEOUT=120
-WARN_TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5 -FGAPOP=6 -FGAPEXT=7
-YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database : PIR 80.*
1: pir1.*
2: pir2.*
3: pir3.*
4: pir4.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	360.5	40.1	627	A40201	artifact-warning s
2	341	37.9	613	A40201	artifact-warning s
C 3	322	35.8	613	C40201	artifact-warning s
C 4	305.5	33.9	627	A40201	artifact-warning s
5	230	25.6	673	A40201	artifact-warning s
C 6	227	25.2	597	A40201	artifact-warning s
.c 7	194.5	21.6	673	A40201	artifact-warning s
8	187.5	20.8	579	A40201	artifact-warning s
9	185	20.6	100	A46010	X-linked retinopat
10	180	20.0	499	S65657	alpha-1C-adrenergi
C 11	178.5	19.8	597	A40201	artifact-warning s
C 12	168.5	18.7	196	2138022	hypothetical prote
13	156	17.3	841	178885	serine/threonine-s
C 14	136	15.1	579	A40201	artifact-warning s

15	128	14.2	407	2	T02670	probable thromboxa
16	126	14.0	369	2	A53959	thromboxane A-2 re
17	123	13.7	79	2	A56194	thromboxane A-2 re
C 18	119.5	13.3	418	2	S41044	chromosomal protei
19	119	13.2	46	2	I54375	gene NP2 protein -
20	118	13.1	39	2	I54374	gene NP2 protein -
21	112	12.4	53	2	A42442	integrin beta-1 ch
22	100	11.1	301	4	B40201	artifact-warning s
23	92	10.2	440	2	A26359	decay-accelerating
24	89	9.9	252	2	T48681	hypothetical prote
25	86	9.6	806	2	F70204	cell division prot
26	86	9.6	806	2	F86600	cell division prot
C 27	85	9.4	357	2	E85049	probable transposo
C 28	83	9.2	301	4	B40201	artifact-warning s
29	83	9.2	522	2	T08711	gamma-adaptin homo
30	82.5	9.2	320	1	A05023	cytochrome c-type
31	81	9.0	347	2	T01452	translation initia
C 32	75.5	8.4	394	2	E90115	26S protease regul
C 33	75.5	8.4	816	1	A59003	phosphoinositide 3
C 34	74.5	8.3	404	2	G82449	amino acid biosynt
C 35	74.5	8.3	601	2	JH0170	octopamine recepto
36	74	8.2	285	2	JC5639	1-acylglycerol-3-p
37	74	8.2	731	2	D36808	hypothetical prote
38	73	8.1	373	1	DKRTH1	3beta-hydroxy-Delt
39	73	8.1	373	1	DERTH2	3beta-hydroxy-Delt
C 40	73	8.1	373	1	DERTHM	3beta-hydroxy-Delt
C 41	73	8.1	571	2	B86150	hypothetical prote
C 42	73	8.1	679	2	S02165	regulatory protein
C 43	72.5	8.1	738	2	A87516	dipeptidyl peptida
C 44	72.5	8.1	962	2	T05845	hypothetical prote
C 45	72.5	8.1	1417	2	F96613	hypothetical prote

ALIGNMENTS

RESULT 1

A40201
artifact-warning sequence (translated ALU class A) - human
C:Species: Homo sapiens (man)
C>Date: 31-Mar-1992 #sequence_revision 11-Aug-1995 #text_change 19-May-2000
C:Accession: A40201
personal communication, 1992
A:Reference number: A40201
A:Accession: A40201
A:Molecule type: DNA
A:Residues: 1-627 <CLA>
A:Cross-references: UNIPARC:UPI000017CEC8
R:Claverie, J.M.
Genomics 12, 838-841, 1992
A:Title: Identifying coding exons by similarity search: Alu-derived and other potential
A:Reference number: A40200; MUID:92241891; PMID:1572661
A:Contents: annotation
C:Comment: This "warning" entry is a conceptual translation in all 6 reading frames of
in-frame stop codons are shown as 'X'.
C:Comment: Any significant similarity of a predicted protein sequence to a portion of

Alignment Scores:

Pred. No.: 1.03e-30 Length: 627
Score: 360.50 Matches: 94
Percent Similarity: 56.5% Conservative: 10
Best Local Similarity: 51.1% Mismatches: 40
Query Match: 40.1% Indels: 40
DB: 4 Gaps: 5

US-10-607-806-1_COPY_5000_5500 (1-501) x A40201 (1-627)

Qy 6 GCTCAGTCTCGCCAGGCTGGAGTGGGTGATCTCAGATCACTACACCTCCA 65
::: |||||:::|||||:::|||||:::|||||:::|||||:::|||||:::|||||
Db 329 SerArgSerValAlaIatysAlaGlyValGlnTrpArgAspLeuGlySerLeuGlnAlaPro 348
Qy 66 TCTCTGGGTTCMAATAATTCTTCTTGCTCAGCCTCTCAAGTAGCTGGGACTTAGCTC 125

[illegible]

RESULT 4

A40201
 artifact-warning sequence (translated ALU class A) - human
 C:Species: Homo sapiens (man)
 C:Date: 31-Mar-1992 #sequence_revision 11-Aug-1995 #text_change 19-May-2000
 C:Accession: A40201
 R:Claverie, J.M.
 personal communication, 1992
 A:Reference number: A40201
 A:Accession: A40201
 A:Molecule type: DNA
 A:Residues: 1-627 <CLA>
 A:Cross-references: UNIPARC:UPI000017CEC8
 R:Claverie, J.M.
 Genomics 12, 838-841, 1992
 A:Title: Identifying coding exons by similarity search: Alu-derived and other potentially
 A:Reference number: A40200; MUID:92241891; PMID:1572661
 A:Contents: annotation
 C:Comment: This "warning" entry is a conceptual translation in all 6 reading frames of c
 in-frame stop codons are shown as 'X'.
 C:Comment: Any significant similarity of a predicted protein sequence to a portion of th

Alignment Scores:	9.73e-25	Length:	627
Pred. No.:	305.50	Matches:	90
Score:	54.3%	Conservative:	12
Percent Similarity:	47.9%	Mismatches:	38
Best Local Similarity:	47.9%	Indels:	48
Query Match:	39.9%	Gaps:	6
DB:	4		

US-10-607-806-1 COPY 5000 5500 (1-501) x A40201 (1-627)

501	QY	GAATCCCGCTCTTTATTAAAAATCAAAAAATTTAGCGGGCGCTGTGGCGAGCGCGCTTT	442
34	Db	GlutThrProSerLeuLeuLysIleGln-LysIleSerArgAla-***CysGlyArgLeu*	53
441	QY	AGTCCACGCTACTGGGGAGGCTGAGCGGGGANGATGGCGTGAACCCAGACAGGTGAGCTT	382
53	Db	**SerGlnLeuLeuGlyArgLeuArgGlnGluAsnGlyValAsnProGlyGlyGlyAlaC	73
381	QY	CGAGTGAAGCCGAGATCGCACCACTGTACTCCAGCGCTTGGCGCAACAGCAAGACTCTGTCT	322
73	Db	ysSerGluProArgSerArgHisCysThrProAlaLeuAlaThrGluAlaGaspSerValS	93
321	QY	CAATTAATAATAGATAATAATAAATGTGTAATAGTAAAGAAACTGGTAATCTGCATTTAGAA	262

93	er	-----GluL 95	
261	AGGGGAGAAATAGA	-----GCAGGATGAGCA	-----G 235
95	ysAsnLysAsnLysLys**Lys*****AlaGlyArgGlyLeuThrProValIleP		115
234	AATCTAAAGGGAGGTCA	CCCGCTGGCCAACTGTTGCAAAACCTGCTCTACCAAAA	175
115	roALeuTriPGluAlaLysAlaGlyGlySerArgGlyGlnGluIleGluThrIleLeuA		135
174	AAACA	-----CACAAAAAATTAGCTGGGCAT	---GGTG 145
135	laThrThrValLysProArgLeuTy*	***LysTy*LYSLysLeuAlaGlyArSerAlaG	155
144	GTGTGTGCCAGCTACTT	GAGAGCTACAAGTCCAGCTACTT	GAGAGGCTGAGGCAAGAGA 85
155	lyAlaCys-Ser	-----ProSerTy*LeuGlyGly***GlyArgArg	168
84	ATTATTGAACCCAGGAGATGGAGGT	TGTAGTGTCTGAGATCAGCCACCCAGCTCCAG	25
169	MetAla***ThrArgGluAlaGluLeuAlaValSerArgAspArgAlaThrAlaLeuGln		188
24	CCTGGCGCACAGGTGAG	7	
189	ProTriArgGlnSerGln	194	

DEPT T C

artifact-warning sequence (translated ALU class F) - human
 C:Species: Homo sapiens (man)
 C:Date: 31-Mar-1992 #sequence_revision 11-Aug-1995 #text_change 19-May-2000
 C:Accession: F40201
 R:Claverie, J.M.
 personal communication, 1992
 A:Reference number: A40201
 A:Accession: F40201
 A:Molecule type: DNA
 A:Residues: 1-673 <CLA>
 A:Cross-references: UNIPARC:UPI0000017CECD
 R:Claverie, J.M.
 Genomics 12, 838-841, 1992
 A:Title: Identifying coding exons by similarity search: Alu-derived and other potential
 A:Reference number: A40200; MUID:92241891; PMID:1572661
 A:Contents: annotation
 C:Comment: this "warning" entry is a conceptual translation in all 6 reading frames of
 in-frame stop codons are shown as 'X'.
 C:Comment: Any significant similarity of a predicted protein sequence to a portion of t

Alignment Scores:		
pred. No.:	1.55e-16	Length: 673
Score:	230.00	Matches: 79
Percent Similarity:	41.2%	Conservative: 17
Best Local Similarity:	33.9%	Mismatches: 44
Query Match:	25.6%	Indels: 93
DB:	4	Gaps: 6

US-10-607-806-1 COPY 5000 5500 (1-501) x F40201 (1-673)

Qy	3	GGGGTCACTCTGTGCCCGAGCTGGAGTGGCGGTGATCTCAGATCACTACAACCT	62
Db	358	GlyProCysSerAspThrProAlaGlyValGlnTrpGlnValLeuAla-HisCysSerLe	377
Qy	63	CAATCTCTGGGTTCAATAATATCTTCTGCTCAGCCTCTCAAGTAGCTGGAGCTGTAG	122
Db	377	uAnLeuLeuGlySerSerAspSerProAlaSerValSerArgValaGly-----	394
Qy	123	CTCTCAAGTAGCTGGCACACACCACCGCCAGCTAAATTTTGTGTGTGTGTGTGTGT	182
Db	395	-----IleThrGlyMetGlyArgHisSerTrpLeuIleTy-----ValPheLeuAl	410
Qy	183	AGAGACAGGTTTCACCATGTGGCCAGGCTGGGTGACCTCCCTTTTAGATTCTCTCAT	242

```
Db      410 eGluThrGlnPheHisValAspGlnAlaGly-----LeuLys 423
      243 CCTGCTCTATTCTTCCCTTTCTAATGCAGTATCCAGTTTCCTTACTATCACAATTTATT 302
      423 sLeuLeuThrSerSer-----AspLeuProSerTrpSerProLysVa 437
      303 ATTA----- 306
      437 lLeuGly**GlnAla**AlaThrThrProSer*****PheValPheGlyGlyPh 457
      307 -----TTCCTATTATTATTAGACAGAGTCTTCTTTGTGCGCAAG 347
      457 ePhePhePhePheAlaLeuPheLeuArg**AlaLeuAlaLeuThrProAr 477
      348 GCTGGAGTACAGTGTGGGATCTCGGCTCACTGCAAGCTCCACCTGCTGGGTTTCACGCCA 407
      477 gLeuGluCysSerGly-LysPheTrpLeuThrAlaAlaSerThrTrpValGlnAlaI 497
      408 TTCTC----- 412
      497 lLeuLeuProLeuSerProVal***LeuGlyLeuGlnAlaTrpAlaAlaIleProGly* 517
      412 ----- 412
      517 **PheMetTyrPhe*****ArgHisSerPheThrMetLeuIleArgLeuValSerAsnS 537
      413 -----CGCCTCAGCTCCCGCTAGCTGGGACTAAGGGCGCTG 452
      537 er***ProGlnValIleCysProPro-GlyLeuProLysCysTrpAspArgArgGlu 556
      453 CCACACGCCCGCTTAATTTTTCGTATTTTA 485
      557 ProProHisProAla*****LeubHeLeu 567

RESULT 6
E40201
artifact-warning sequence (translated ALU class E) - human
C:Species: Homo sapiens (man)
C:Date: 31-Mar-1992 #sequence_revision 11-Aug-1995 #text_change 19-May-2000
C:Accession: E40201
R:Claverie, J.M.
personal communication, 1992
A:Reference number: A40201
A:Accession: E40201
A:Molecule type: DNA
A:Residues: 1-597 <CLA>
A:Cross-references: UNIPARC:UPI0000017CECC
R:Claverie, J.M.
Genomics 12, 838-841, 1992
A:Title: Identifying coding exons by similarity search: Alu-derived and other potentiall
A:Reference number: A40200; MUID:92241891; PMID:1572661
A:Contents: annotation
C:Comment: This "warning" entry is a conceptual translation in all 6 reading frames of t
in-frame stop codons are shown as 'X'.
C:Comment: Any significant similarity of a predicted protein sequence to a portion of th

Alignment Scores:
Pred. No.: 3,27e-16 Length: 597
Score: 227.00 Matches: 73
Percent Similarity: 50.6% Conservative: 14
Best Local Similarity: 42.4% Mismatches: 56
Query Match: 25.2% Indels: 29
DB: 4 Gaps: 10

US-10-607-806-1_COPY_5000_5500 (1-501) x E40201 (1-597)

QY 500 AAACCCCGCTCTTTTAAATAACAAAAAATTAGCGGGCGTGGTGGCAGCGCCTTTA 441
Db 36 LysProHisLeuTyr***LysTyr---LysIleSerTrpAlaTrpArgMetProVal 54
QY 440 GTCCACGACTACTGGGGAGGCTGAGGGGGAGAA---TGGCGTGAACCCAGCAGGTGGAGC 384
```

```
Db      55 lIleProAlaThrArgGluAlaGluThrGlyGluAsnCysLeuAsnProGly-----Ser 72
      383 TTCAGTGGAGCCAGATCGCACCTGTACTCCAGCTTGGCGACAAAGCAAGCACTCTGT 324
      73 LysValCysGlyIleValProSerTyrSerGlyLeuGlyAsnLysSerLysThrPro 92
      323 CTCATAATAATAAGAAATAATAAT---AAATGTGATAAGTAAGAAACTGGATACTGCAT 267
      93 SerGlnLysLysLys*****AlaArgCys-----GlySerSerCysLeu** 109
      266 TAGAAAGGGGAAGAATAAGACAGGATGAGGAGAATCTAAAGGGAGGT----- 219
      110 SerGlnArgPheGlySerSerArg-----ArgGlyGlySerProGluVal 124
      218 -----CACCCAGCTCGCCAAACATGTGTGAAAACCTCTCTCTACCAAAAAACACA 168
      125 GlySerLeuArgProAla***ProThrTrpArgAsnProIleSerThr-LysAsnIle-- 143
      167 CAAAAAATTAGCTGGGCATGTGTGTGCCAGCTACTTTGAGAGCTACAAGTCCCAAGCT 108
      144 ---LysLeuAlaGlyArgGlyGlyAlaCys-----Leu**SerGlnLe 157
      107 ACTTGACAGGCTGAGGCAAGAGAAATTATTGAACCCAGAGATGAGGTTGTAGTATCT 48
      157 uLeuGlyArgLeuArgGlnGluArgIleAla***ThrGlnGluAlaLysPheAla--Va 176
      47 GAGATCAGCCACCGCCTCAGCTCGAGTGGCGACA 14
      176 lArgLeuCysHisArgThrProAlaTrpAlaThr 187

RESULT 7
F40201
artifact-warning sequence (translated ALU class F) - human
C:Species: Homo sapiens (man)
C:Date: 31-Mar-1992 #sequence_revision 11-Aug-1995 #text_change 19-May-2000
C:Accession: F40201
R:Claverie, J.M.
personal communication, 1992
A:Reference number: A40201
A:Accession: F40201
A:Molecule type: DNA
A:Residues: 1-673 <CLA>
A:Cross-references: UNIPARC:UPI0000017CECD
R:Claverie, J.M.
Genomics 12, 838-841, 1992
A:Title: Identifying coding exons by similarity search: Alu-derived and other potentiall
A:Reference number: A40200; MUID:92241891; PMID:1572661
A:Contents: annotation
C:Comment: This "warning" entry is a conceptual translation in all 6 reading frames of c
in-frame stop codons are shown as 'X'.
C:Comment: Any significant similarity of a predicted protein sequence to a portion of th

Alignment Scores:
Pred. No.: 1,11e-12 Length: 673
Score: 194.50 Matches: 71
Percent Similarity: 46.0% Conservative: 16
Best Local Similarity: 37.6% Mismatches: 57
Query Match: 21.6% Indels: 45
DB: 4 Gaps: 4

US-10-607-806-1_COPY_5000_5500 (1-501) x F40201 (1-673)

QY 501 GAAACCCCGCTCTTTTAAATAACAAAAAATTAGCGGGCGTGGTGGCAGCGCCTTT 442
Db 35 GluThrValSerLeuLeuLysIleHis--LysLeuAlaArgAsnGlyGlyProCysLeu* 54
QY 441 AGTCCAGCTACTGGGGAGGCTGAGCGGGAGAAATGGCGTGAACCCAGCAGGTGGAGCTT 382
Db 54 **SerGlnLeuHisGlyArgGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGly**GlyC 74
QY 381 GCAGTGGAGCGAGATCGCACCACTGTACTCCAGCCTTGGCGACAAAGCAAGACTCTGTCT 322
```

```
Db 74 ySerGluProLys-LeuAlaThrAlaLeuGlnProGlyCysGlnSerLysGlyLeuSer 93
QY 321 CAATAATAATAAGAAATAATAATAAGTATAGTAAGTAAGAACTGGATGACTGCATTAGAA 262
Db 94 Gln-----LysGlnLysGlnSerLysLysLys 102
QY 261 AGGGGAAGAATAGACGAGATGAGGAGAACTTAAAGGG----- 223
Db 103 LysLysLysThrProLysAsnLys*****AlaGlyCysGlyGlySerArgLeuSer 122
QY 222 -----AGTCACCCAGC 211
Db 123 SerGlnHisPheGlyArgProGlyGlyGlnIleThr****GlyGlnGluPheGluThrSer 142
QY 210 CTGGCCCAACATGCTGAAAACCTGTCTTACCAAAAAAACAACACAAAAAATTAGCTGGC 151
Db 143 LeuIleAsnMetValLysLeuCysLeuTy****Lys-----TyIleAsn***ProGly 160
QY 150 ATGGTGGTGTGGCCAGCTACTTGGAGAGTACAAGTCCAGCTACTTGGAGGGCTGAGGC 91
Db 161 MetAlaAla-----HisAlaCysAsnProSerTyThrGlyAspArgGly 175
QY 90 AAGAGAATTATTGAACCCAGGATGGAGGTGTAGTCACTGAGATCACGCCACCCCA 31
Db 176 ArgArgIleAla***ThrGlnGluValGluAlaAlaVal-SerGlnAsnLeuProLeuHi 195
QY 30 CTCCAGCCTGGCGCACAGA 12
Db 195 sSerSerArgGlyValArg 201
```

RESULT 8

```
D40201
A:Title: artifact-warning sequence (translated ALU class D) - human
C:Species: Homo sapiens (man)
C>Date: 31-Mar-1992 #sequence_revision 11-Aug-1995 #text_change 19-May-2000
C:Accession: D40201
R:Claverie, J.M.
A:Reference number: A40201
A:Accession: D40201
A:Molecule type: DNA
A:Residues: 1-579 <CLA>
A:Cross-references: UNIPARC:UPI000017CECB
R:Claverie, J.M.
Genomics 12, 838-841, 1992
A:Title: Identifying coding exons by similarity search: Alu-derived and other potential
A:Reference number: A40200; MUID:92241891; PMID:1572661
A:Contents: annotation
C:Comment: This "warning" entry is a conceptual translation in all 6 reading frames of c
in-frame stop codons are shown as 'X'.
C:Comment: Any significant similarity of a predicted protein sequence to a portion of th
```

Alignment Scores:

```
Pred. No.: 6,38e-12 Length: 579
Score: 187.50 Matches: 74
Percent Similarity: 50.6% Conservative: 16
Best Local Similarity: 41.6% Mismatches: 57
Query Match: 20.8% Indels: 31
DB: 4 Gaps: 5
```

US-10-607-806-1_COPY_5000_5500 (1-501) x D40201 (1-579)

```
QY 3 GGGGCTCACTCTGTCGCCAGCTGGAGTGGGTGGCGTGATCTCAGATCACTACAACCT 62
Db 392 GlySerTrpTyrValAlaGlnAlaGlyValGlnArgCysAsnHisGlySerLeuSerLeu 411
QY 63 CCATCTCTCGGTTCAATAATCTCTTGCTCTCAGCTCTCAAGTAGCTGGAGCTTAG 122
Db 412 AspPheProGly--GlnValIleLeuLeuProGlnProPheGlu***LeuGlyLeu---- 429
QY 123 CTCTCAAGTAGCTGGCACACACCACCATGCCAGCTAATTTTGTGTGTTTTTTTGGT 182
```

```
Db 430 -----GlnAlaTyzAlaThrArg-SerGly***PheLeuTyzLeuPhePhe-Va 445
QY 183 AGACAGAGTGTTCACCATGTTGCCAGCTGGGTGGACCTCCCTTTTATGATTCCTCAT 242
Db 445 lGluValGlyPheArgHisValAlaGlnAla-----ValLeuGluLe 459
QY 243 CTGTCTTATTTCTTCCCTTTCTAATGCAGTATCCAGTTTCTTACTTTA----- 291
Db 459 uLeuSerTrp-----SerIleHisLeuProTrpLeuLeuLysCysTr 473
QY 292 -----TCACATTATTATTATTATTATTATTATTAGACAGAGCTTTGCTTTGTC 341
Db 473 pAspTyzArgArgAlaAlaMetLeuGly*****PheArgGlnGlyLeuGlyMetLe 493
QY 342 GCCAAGGCTGAGTACAGTGTGCGATCTCGCTCTCACTGCAAGCTCCACCTGCTGGTTC 401
Db 493 uProArgLeuGluTyzArgGlyAlaIleMetAlaHis***AlaLeu-ThrSerGlnVal 513
QY 402 ACGCCATTCTCCCGCTCAGCTCCCGAGTAGCTGGGACTAAAGCGCGCTGCCACCCAGC 461
Db 513 ys***SerSerCysPro-SerLeuLeuSerSerTrpAspTyzArgLeuMetLeuProAsp 532
QY 462 CCCGCTAATTTTTTTGTGTA---TTTTTAATAAAGACGGGTTT 500
Db 533 LeuAlaAsnPheCysIlePhePheLeu***ArgTrpGlyPhe 546
```

RESULT 9

```
A46010
A:Title: X-linked retinopathy protein (C-terminal, clone XEH.8c) - human (fragment)
C:Species: Homo sapiens (man)
C>Date: 21-Sep-1993 #sequence_revision 18-Nov-1994 #text_change 09-Jul-2004
C:Accession: A46010
R:Wong, P.; MacDonald, I.M.; Sood, R.; Smith, C.; Pilon, R.; Tenniswood, M.
Genomics 15, 467-471, 1993
A:Title: Identification and partial characterization of a candidate gene for X-linked
A:Reference number: A46010; MUID:93224131; PMID:8468040
A:Accession: A46010
A:Status: preliminary
A:Molecule type: nucleic acid
A:Residues: 1-100 <WON>
A:Cross-references: UNIPROT:O07826; UNIPARC:UPI000006B23F; GB:S58722; NID:9299470; PID:
A>Note: sequence extracted from NCBI backbone (NCBIN:129339, NCBIP:129340)
```

Alignment Scores:

```
Pred. No.: 1.1e-11 Length: 100
Score: 185.00 Matches: 42
Percent Similarity: 77.0% Conservative: 5
Best Local Similarity: 68.9% Mismatches: 11
Query Match: 20.6% Indels: 3
DB: 2 Gaps: 0
```

US-10-607-806-1_COPY_5000_5500 (1-501) x A46010 (1-100)

```
QY 321 GAGACAGAGTCTTGCTTTGTCGCAAGCTGGAGTACAGTGGTGGATCTCGGCTCACTG 380
Db 7 GluThrGluSerCysSerValAlaGluAlaGlyValGlnTrpCysAspLeuGlySerLeu 26
QY 381 CAAGCTCCACTGCTGGTTTCAGCCCATTTCTCCGCTCAGCTCTCCACAGTAGCTGGAC 440
Db 27 LysSerProProGlySerSerAspSerProAlaSerAlaSerArgValAlaGlyI 46
QY 441 TAAAGCGCTGCCACACCGCCCGCTAATTTTGTGTTTATTAAGACGGG 497
Db 46 eThrGlyMetHisHisThrGlnLeuIle--PheValPheLeuValGluThrGly 64
```

RESULT 10

```
S65657
A:Title: alpha-1C-adrenergic receptor splice form 2 - human
N:Alternate names: alpha-1C-adrenoceptor isoform 2
C:Species: Homo sapiens (man)
C>Date: 22-Nov-1996 #sequence_revision 22-Nov-1996 #text_change 09-Jul-2004
C:Accession: S65657; S65655
```

R: Tanaka, T.
submitted to the EMBL Data Library, July 1994
A: Reference number: S65656
A: Accession: S65657
A: Molecule type: mRNA
A: Residues: 1-499 <TAN>
A: Cross-references: UNIPROT: Q13675; UNIPARC: UPI000004E75E; EMBL: D32202; NID: G927208; PID: R.Hirasawa, A.; Shibata, K.; Horie, K.; Takei, Y.; Obika, K.; Tanaka, T.; Muramoto, N.; FEBS Lett. 363, 256-260, 1995
A: Title: Cloning, functional expression and tissue distribution of human alpha (1C)-adren
A: Reference number: S65654; MUID: 95255557; PMID: 7737411
A: Accession: S65655
A: Molecule type: mRNA
A: Residues: 424-499 <HR>
A: Cross-references: UNIPARC: UPI00001778AB; EMBL: D32202
C: Genes: GDB: ADRA1C; ADRA1L1
A: Gene: GDB: ADRA1C; ADRA1L1
A: Cross-references: GDB: 128088; OMIM: 104221
A: Map position: 8p21-8p11.2
C: Superfamily: vertebrate rhodopsin
C: Keywords: alternative splicing; neurotransmitter receptor

Alignment Scores:
Pred. No.: 4.14e-11 Length: 499
Score: 180.00 Matches: 40
Percent Similarity: 66.7% Conservative: 2
Best Local Similarity: 63.5% Mismatches: 19
Query Match: 20.0% Indels: 2
DB: 2 Gaps: 1

US-10-607-806-1_COPY_5000_5500 (1-501) x S65657 (1-499)

QY 22 AGGCTGGAGTGGGGTGGCTGATCTCAGATCACTACAACTCCATCTCTCTGGTTCAAAAT 81
DB 431 ArgLeuGluCysSerGlyMetileLeuAlaHisCysAsnLeuArgLeuProGlySerArg 450
QY 82 AATTCTTTCCTCAGCTCTCAAGTAGCTGGAGCTTGTAGCTCTCAAGTAGCTGGCACA 141
DB 451 AspSerProAlaSerAlaSerGlnAlaGlyThrGlyAspValProGlyArg 470
QY 142 CACCACCATGCCAGCTAATTTTTGTGTGTTTTTTTGGTAGACAGAGTTTTCCACCAT 201
DB 471 ArgHisGlnAlaGlnLeullePhe-----ValPheLeuValGluThrGlyPheHis 488
QY 202 GTTGCCAG 210
DB 489 ValGlyGln 491

RESULT 11
E40201
artifical-warming sequence (translated ALU class E) - human
C: Species: Homo sapiens (man)
C: Date: 31-Mar-1992 #sequence_revision 11-Aug-1995 #text_change 19-May-2000
C: Accession: E40201
R: Claverie, J.M.
personal communication, 1992
A: Reference number: A40201
A: Accession: E40201
A: Molecule type: DNA
A: Residues: 1-597 <CLA>
A: Cross-references: UNIPARC: UPI000017CECC
R: Claverie, J.M.
Genomics 12, 838-841, 1992
A: Title: Identifying coding exons by similarity search: Alu-derived and other potential
A: Reference number: A40200; MUID: 92241891; PMID: 1572661
C: Contents: annotation
C: Comment: This "warning" entry is a conceptual translation in all 6 reading frames of c
in-frame stop codons are shown as 'X'.
C: Comment: Any significant similarity of a predicted protein sequence to a portion of th

Alignment Scores:
Pred. No.: 6.07e-11 Length: 597

Score: 178.50 Matches: 62
Percent Similarity: 46.2% Conservative: 16
Best Local Similarity: 36.7% Mismatches: 62
Query Match: 19.8% Indels: 29
DB: 4 Gaps: 9

US-10-607-806-1_COPY_5000_5500 (1-501) x E40201 (1-597)

QY 31 TGGGTGGG-----GTGATCTCAGATCACTACAACTCCATCTCTCTGGTTCAAAAT 81
DB 412 CysProGlyArgSerThrMetAlaGlnSerHisArgLysLeuCysPheLeuGlySerSer 431
QY 82 AATTCT---CTTGCTCAGCTCTCAAGTAGCTGGAGCTTGTAGCTCTCAAGTAGCTGGC 138
DB 432 AsnSerLeuLeuSer-GlnProGlu***LeuGlyLeuGlnAlaCys----- 447
QY 139 ACACACCATGCCAGCTAAATTTTTGTGTGTTTTTTTGGTAGACAGAGTTTTTCAC 198
DB 448 ---AlaThrThrProSer***PheTyrIle-----Phe-SerArgAspGlyValSerP 464
QY 199 CATGTGGCCAGGCTGG-----GTGACCTCCCTTTTGTAGATTCTCTCATCTGCTCTAT 252
DB 464 roCysTrpSerSerTrpSerGlnThrProAsnLeuArg***SerAlaProPro***Thr- 483
QY 253 TCTTCCCTTTCTAATGCAGTATCCAGTTCTTCTTACTTATCATCATTTTATTATTCTTA 312
DB 484 ----ProLysAlaLeuGlyLeuGlnAla***AlaThrThrProGly*****PheP 502
QY 313 TTATTATTGAGACAGAGTCTTGTGTCGCAAGGCTGGAGTACAGTGGTGCATCTCG 372
DB 502 hePheLeuArgArgSerPheAlaLeuValAlaGlnAlaGlyValArgTrpHisAsnLeu- 521
QY 373 GCTCAGTCAAGCTCCACCTGCTGGTTCACGCCATTTC-----TCCCGCC 417
DB 522 -----ThrAlaAsnPheAlaSerTrpValGlnAlaLeuLeuSerCysL 536
QY 418 TCAGCTCCCGAGTGGAGCTAAAGGGCGCTGCCACACCGCCCTAAATTTTTTTG 477
DB 536 euSerLeuProSerSerTrpAspTrpArgHisAlaProProArgProAlaAsn---PheI 555
QY 478 TATTTTAAATAAGACGGGGTTT 500
DB 555 lePheLeuValGluMetGlyPhe 562

RESULT 12
I38022
hypotheical protein - human
C: Species: Homo sapiens (man)
C: Date: 01-Nov-1996 #sequence_revision 01-Nov-1996 #text_change 05-Nov-1999
C: Accession: I38022
R: Yang, S.S.; Zhang, K.; Vieira, W.; Taub, J.V.; Zellstra-Ryalls, J.H.; Somerville, R.L.
Cancer Res. 50, 5658-5667, 1990
A: Title: A human hepatocellular carcinoma 3.0-kilobase DNA sequence transforms both rat
A: Reference number: I38021
A: Accession: I38022
A: Status: preliminary; translated from GB/EMBL/DBJ
A: Molecule type: DNA
A: Residues: 1-196 <RES>
A: Cross-references: UNIPARC: UPI000011EIFB; EMBL: X55777; NID: G288143; PIDN: CAA39297.1; PT

Alignment Scores:
Pred. No.: 7.05e-10 Length: 196
Score: 168.50 Matches: 51
Percent Similarity: 56.0% Conservative: 14
Best Local Similarity: 44.0% Mismatches: 38
Query Match: 18.7% Indels: 13
DB: 2 Gaps: 3

US-10-607-806-1_COPY_5000_5500 (1-501) x I38022 (1-196)

QY 495 CCGTCTTTTAAATAACAAAAAATAGCGGGCGTGGTGGCAGGCCCTTTAGTCCC 436
DB 19 ProSerLeuLeuLysIleGln--GluLeuAlaGlyHisGlyGlyArgCysLeuGlnSerG 38

A;Description: Sequence analysis of a 3.5 Mb contig in human 19p13.3 containing a serine
A;Reference number: Z14696
A;Accession: T02670
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-407 <LAM>
A;Cross-references: UNIPROT:O75228; UNIPARC:UPI00000711C4; EMBL:AC005175; NID:G3253116;
C;Superfamily: prostaglandin E receptor EPI

Alignment Scores:
Pred. No.: 1.83e-05 Length: 407
Score: 128.00 Matches: 30
Percent Similarity: 60.4% Conservative: 2
Best Local Similarity: 56.6% Mismatches: 21
Query Match: 14.2% Indels: 0
DB: 2 Gaps: 0

US-10-607-806-1_COPY_5000_5500 (1-501) x T02670 (1-407)

QY	322	AGACAGAGCTTTGCTTTGTCGCAAGGCTGGAGTACAGTGTGCGATCTCGGCTCACTGC	381
		:	
Db	328	ArgArgSerLeuThrLeuTriProSerLeuGluTyrSerGlyThrIleSerAlaHisCys	347
QY	382	AAGCTCCACCTGCTGGTTTCAGGCCATTCTCCGCTCAGCCTCCCACTAGCTGGGACT	441
Db	348	AsnLeuArgLeuProGlySerSerAspSerArgAlaSerArgAlaAlaGlyIle	367
QY	442	AAAGGGGCTGCGCACGCGCGCTAATTTTGTAT	480
Db	368	ThrGlyValSerHisCysAlaArgProCysMetLeuPhe	380

Search completed: February 8, 2006, 10:03:35
Job time : 24.412 secs

GenCore version 5.1.7
Copyright (c) 1993 - 2006 Bioacceleration Ltd.

OM nucleic - protein search, using frame_plus_n2p model

Run on: February 8, 2006, 09:18:47 ; Search time 24.3157 Seconds
(without alignments)
2907.338 Million cell updates/sec

Title: US-10-607-806-1_COPY_5000_5500
Perfect score: 900
Sequence: 1 acggggctcactctgtgcc.....tttaataaagcggggttc 501

Scoring table:
BLOSUMP62
Xgapop 10.0 , Xgapext 0.5
Ygapop 10.0 , Ygapext 0.5
Fgapop 6.0 , Fgapext 7.0
Delop 6.0 , Delext 7.0

Searched: 2166443 seqs, 705528306 residues

Total number of hits satisfying chosen parameters: 4332886

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Command line parameters:
-MODEL=frame+n2p.model -DEV=xlp
-O=/abs/ABSSWEB spool/US10607806/runat_08022006_091819_26843/app_query.fasta_1
-DB=UniProt -QFWT=fastan -SUFFIX=n2p.rup -MINMATCH=0.1 -LOOPCL=0 -LOOPEXT=0
-UNITS=bits -START=1 -END=1 -MATRIX=blosump62 -TRANS=human40.cdi -LIST=45
-DOCALLIGN=200 -THR SCORES=pct -THR MAX=100 -THR MIN=0 -ALIGN=15 -MODE=LOCAL
-OUTFWT=ptc -NORM=ext -HEAPSIZ=500 -MINLEN=0 -MAXLEN=200000000 -HOST=abs02p
-USER=US10607806 @CGN.1.1.694 @runat_08022006_091819_26843 -NCPU=6 -ICPU=3
-NO_MMAP -NEG SCORES=0 -WAIT -DSBLOCK=100 -LONGLOG -DEV TIMEOUT=120
-WARN TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5 -FGAPOP=6 -FGAPEXT=7
-YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database : UniProt.05.80.*
1: uniprot.spot.*
2: uniprot.trembl.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES						
Result No.	Score	Query Match %	Length	ID	Description	
1	391	43.4	170	Q8NA13_HUMAN	Q8NA13 homo sapien	
2	386	42.9	585	1 ALU5_HUMAN	P39192 homo sapien	
3	366	40.7	587	1 ALU2_HUMAN	P39189 homo sapien	
4	363.5	40.4	591	1 ALU8_HUMAN	P39195 homo sapien	
5	362	40.2	587	1 ALU1_HUMAN	P39189 homo sapien	
6	359	39.9	593	1 ALU7_HUMAN	P39194 homo sapien	
7	358.5	39.8	179	2 Q6ZS15_HUMAN	Q6ZS15 homo sapien	
8	351.5	39.1	591	1 ALU8_HUMAN	P39195 homo sapien	
9	350	38.9	587	1 ALU3_HUMAN	P39190 homo sapien	
10	344.5	38.3	158	2 Q8NA19_HUMAN	Q8NA19 homo sapien	
11	342.5	38.1	591	1 ALU1_HUMAN	P39189 homo sapien	
12	339	37.7	585	1 ALU5_HUMAN	P39192 homo sapien	
13	337	37.4	593	1 ALU7_HUMAN	P39194 homo sapien	
14	333	37.0	593	1 ALU6_HUMAN	P39193 homo sapien	
15	331	36.8	587	1 ALU3_HUMAN	P39190 homo sapien	
16	328	36.4	593	1 ALU6_HUMAN	P39193 homo sapien	

c	17	323.5	35.9	591	1	ALU1_HUMAN	P39188 homo sapien
	18	314.5	34.9	375	2	Q60448_HUMAN	Q60448 homo sapien
	19	312	34.7	603	1	ALU4_HUMAN	P39191 homo sapien
	20	311	34.6	239	2	Q5ZWA9_HUMAN	Q5ZWA9 homo sapien
	21	295.5	32.8	195	2	Q5ZUN2_HUMAN	Q5ZUN2 homo sapien
	22	293	32.6	200	2	Q5ZUK3_HUMAN	Q5ZUK3 homo sapien
	23	288	32.0	334	2	Q5ZR97_HUMAN	Q5ZR97 homo sapien
	24	284.5	31.6	143	2	Q6ZSJ7_HUMAN	Q6ZSJ7 homo sapien
	25	280	31.1	176	2	Q6ZR92_HUMAN	Q6ZR92 homo sapien
c	26	277.5	30.8	603	1	ALU4_HUMAN	P39191 homo sapien
	27	273	30.3	143	2	Q5ZS55_HUMAN	Q5ZS55 homo sapien
	28	267.5	29.7	241	2	Q6ZTY3_HUMAN	Q6ZTY3 homo sapien
	29	264	29.3	155	2	Q6ZSW8_HUMAN	Q6ZSW8 homo sapien
	30	262.5	29.2	170	2	Q6ZV14_HUMAN	Q6ZV14 homo sapien
	31	262	29.1	162	2	Q9H5D5_HUMAN	Q9H5D5 homo sapien
	32	260.5	28.9	124	2	Q6ZSG7_HUMAN	Q6ZSG7 homo sapien
	33	260.5	28.9	129	2	Q6ZUW8_HUMAN	Q6ZUW8 homo sapien
	34	260.5	28.9	176	2	Q6ZTU6_HUMAN	Q6ZTU6 homo sapien
	35	255.5	28.4	358	2	Q6ZRP1_HUMAN	Q6ZRP1 homo sapien
	36	250.5	27.8	131	2	Q6ZTD3_HUMAN	Q6ZTD3 homo sapien
	37	246.5	27.4	350	2	Q96NA4_HUMAN	Q96NA4 homo sapien
	38	244.5	27.2	160	2	Q6ZNP5_HUMAN	Q6ZNP5 homo sapien
	39	243.5	27.1	156	2	Q8NBH4_HUMAN	Q8NBH4 homo sapien
	40	243	27.0	134	2	Q6ZSG8_HUMAN	Q6ZSG8 homo sapien
	41	238.5	26.5	153	2	Q6ZNX9_HUMAN	Q6ZNX9 homo sapien
c	42	236	26.2	135	2	Q6ZRW7_HUMAN	Q6ZRW7 homo sapien
	43	235.5	26.2	120	2	Q9H6G8_HUMAN	Q9H6G8 homo sapien
	44	234.5	26.1	120	2	Q8N2C8_HUMAN	Q8N2C8 homo sapien
	45	234.5	26.1	150	2	Q6ZR91_HUMAN	Q6ZR91 homo sapien

ALIGNMENTS

RESULT 1
Q8NA13_HUMAN
ID Q8NA13_HUMAN PRELIMINARY; PRT; 170 AA.
AC Q8NA13;
DT 01-OCT-2002 (TrEMBLrel. 22, Created)
DT 01-OCT-2002 (TrEMBLrel. 22, Last sequence update)
DT 01-OCT-2002 (TrEMBLrel. 22, Last annotation update)
DE Hypothetical protein FLJ35322.
OS Homo sapiens (human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini; Hominidae;
OC Homo.
OX NCBI_TaxID=9606;
RN [1]
RP NUCLEOTIDE SEQUENCE.
RC TISSUE=Prostate;
RX PubMed=14702039; DOI=10.1038/ng1285;
RA Ota T., Suzuki Y., Nishikawa T., Oeueki T., Sugiyama T., Irie R.,
RA Wakamatsu A., Hayashi K., Sato H., Nagai K., Kimura K., Makita H.,
RA Sekine M., Oobayashi M., Nishi T., Shibahara T., Tanaka T., Ishii S.,
RA Yamamoto J.-I., Saito K., Kawai Y., Isono Y., Nakamura Y.,
RA Nagahara K., Murakami K., Yasuda T., Iwayanagi T., Wagatsuma M.,
RA Shiratori A., Sudo H., Hosoiri T., Kaku Y., Kodaira H., Kondo H.,
RA Sugawara M., Takahashi M., Kanda K., Yokoi T., Furuya T., Kikkawa E.,
RA Omura Y., Abe K., Kamiyama K., Katsuta N., Sato K., Tanikawa M.,
RA Yamazaki M., Ninomiya K., Ishibashi T., Yamashita H., Murakawa K.,
RA Fujimori K., Tanai H., Kimata M., Watanabe M., Hirakawa S., Chiba Y.,
RA Ishida S., Ono Y., Takiguchi S., Watanabe S., Yosida M., Hotuta T.,
RA Kusano J., Kanehori K., Takahashi-Fujii A., Hara H., Tanase T.-O.,
RA Nomura Y., Togiya S., Komai F., Hara R., Takeuchi K., Arita M.,
RA Inose N., Musashino K., Yuuki H., Oshima A., Sasaki N., Aotsuka S.,
RA Yoshikawa Y., Matsunawa H., Ichihara T., Shiohara N., Sano S.,
RA Moriya S., Momiya H., Satoh N., Takami S., Terashima Y., Suzuki O.,
RA Nakagawa S., Senoh A., Mizoguchi H., Goto Y., Shimizu F., Wakebe H.,
RA Hishigaki S., Watanabe T., Sugiyama A., Takemoto M., Kawakami B.,
RA Yamazaki M., Watanabe K., Kumagai A., Itakura S., Fukuzumi Y.,
RA Fujimori Y., Komiyama M., Tashiro H., Tanigami A., Fujiwara T.,
RA Ono T., Yamada K., Fujii Y., Ozaki K., Hirao M., Ohmori Y.,
RA Kawabata A., Hikiji T., Kobatake N., Inagaki H., Ikema Y., Okamoto S.,
RA Okitani R., Kawakami T., Noguchi S., Itoh T., Shigeta K., Senba T.,

Query Match:	42.9%	Indels:	20
DB:	1	Gaps:	6
US-10-607-806-1_COPY_5000_5500 (1-501) x ALU5_HUMAN (1-585)			
QY	8	TCACCTGTGCGCCAGGCTGGAGTGGGCTGATCTCAGATCAGTACAGTCCATC	67
DB	399	SerLeuGlyArg-GlnAlaGlyValGlnTrpArgAspLeuGlySerLeuGlnProPro	418
QY	68	TCCTGGGTTCAATAATTCCTTGGCTCAGCTCTCAAGTACGCTGGGACTGTAGCTCTC	127
DB	418	oProGlyPheLeuArgPheSerCysLeuSerLeuProSerSerTrpAspTyr	435
QY	128	AGTAGCTGGCACACACACACACACATGCCAGCTGAGTATTTTGTGTGTTTGTGAGAGA	187
DB	436	-----ArgArgAlaProProArgProAlaAsnPhe---CysIle---PheSerArgAs	451
QY	188	CAGGTTTTCACATGTTGGCCAGGCTGGGTGACCTCCCTT-----TTAGATTCTCC	238
DB	451	pGlyValSerProCysTrpProGlyTrpSerArgSerLeuAspLeuValIleArgProPr	471
QY	239	TCATCTGCTCTATTCTTCCCTTCTTAATGACAGTATCCAGTCTTCTTACTTATCATATT	298
DB	471	oArgPro-----ProLysValLeu-----GlyLeuGlnIle***AlaThrAl	485
QY	299	TATTATTATTCTTATTATTATTTAGACACAGAGCTTGTCTTGTCCGCAAGCTGGAGTACA	358
DB	485	aProGly*****PhePheGluThrGluSerArgSerValAla-ArgLeuGluCys	505
QY	359	GTGTGCGATCTCGGCTACTGCAAGCTCCAGCTGCTGGTTCAGGCGATTTCCCGCCT	418
DB	505	erGlyAlaIleSerAlaHisCysAsnLeuArgLeuProGlySerSerAspSerProAlaS	525
QY	419	CAGCTCCCGAGTGTGGAGCTAAAGGCGCTGCACACGCGCCGCTAATTTTGT	478
DB	525	erAlaSerArgValAlaGlyThrThrGlyAlaArgHisAlaGlnLeuIle--PheVa	544
QY	479	ATTTTTAATAAGAGGGGTTT 500	
DB	544	lPheLeuValGluThrGlyPhe 551	
RESULT 3			
ID	ALU2_HUMAN	STANDARD;	PRT; 587 AA.
AC	P39189;		
DT	01-FEB-1995 (Rel. 31, Created)		
DT	01-FEB-1995 (Rel. 31, Last sequence update)		
DE	10-MAY-2005 (Rel. 47, Last annotation update)		
DE	Alu subfamily SB sequence contamination warning entry.		
OS	Homo sapiens (Human)		
OC	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;		
OC	Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini; Homiidae;		
OC	Homo.		
OX	NCBI_TaxID=9606;		
RN	[1]		
RP	NUCLEOTIDE SEQUENCE.		
RP	MEDLINE=95021758; PubMed=7935834; DOI=10.1038/371752a0;		
RA	Claverie J.-M., Makalowski W.;		
RA	"Alu alert.";		
RL	Nature 371:752-752(1994).		
RL	[2]		
RP	CONCEPT.		
RP	MEDLINE=92241891; PubMed=1572661;		
RA	Claverie J.-M.;		
RA	"Identifying coding exons by similarity search: alu-derived and other		
RT	potentially misleading protein sequences.";		
RL	Genomics 12:838-841(1992).		
RL	[3]		
RP	ALU FAMILIES CLASSIFICATION.		
RP	MEDLINE=88333009; PubMed=3138422;		
RA	Quentin Y.;		
RA	"The Alu family developed through successive waves of fixation closely		
RT	connected with primate lineage history.";		


```
QY 187 ACAGGTTTTCACATTTGGCCAGGCTGGTG----- 218
DB 357 ThrGlyPheHisValGlyGlnAlaGlyLeuGluLeuThrSerGlyAspLeuPro 376
QY 218 ----- 218
DB 377 ProLysValLeuGlyLeuGlnAla***AlaThrAlaProGlyLeuArgPro***** 396
QY 218 ----- 218
DB 397 PhePhe**AspGlyValSerLeuCysArgProGlyTrpSerAlaValAlaArgSerArg 416
QY 219 -----ACTCC-----CTTTTAGATTCTCT----- 239
DB 417 LeuThrAlaThrSerAlaSerArgValGlnAlaIleLeuLeuLeuProGlnProProGlu*** 436
QY 240 -----CATCTGCTCTATTCTTCCCTTTCTAATGACGATAT 275
DB 437 LeuGlyLeuGlnAlaArgAlaThrThrProGly***PheLeuTyPhe*****ArgArg 456
QY 276 CCAGTTTCTCTTATCATCATTTATT-----ATTATCTTTATT--- 314
DB 457 GlyPheThrMetLeuAlaArgLeuValSerAsnSer***ProGlnValIlePheLeuPro 476
QY 315 -----ATTT-----ATT 317
DB 477 LysCysTrpAspTyArgArgGluProProArgProAlaSerAlaArg*****Phe 496
QY 318 ATTGACAGAGCTCTGTTTCTCCAGGCTGAGTACAGTGTGGGATCTGGCTCA 377
DB 497 PheGluThrGluSerArgSerValAlaGlnAlaGlyValGlnTrpArgAspLeuGlySer 516
QY 378 CTGCAAGCTCCACCTGCTGGGTTTCAGCCATTCTCCGCTCAGCTCCAGTACGCTGG 437
DB 517 LeuGlnProProProGlyPheLysArgPheSerCysLeuSerLeuProSerSerTrp 536
QY 438 GACTAAGCGCTCCACCCAGCCCGCTAATTTTTTTTGTATTTTAAATAAGACGGGG 497
DB 537 AspTyArgArgAlaProProArgProAlaAsn--PheCysIlePheSerArgAspGlyV 556
QY 498 TT 499
DB 556 al 556

RESULT 5
ALU2_HUMAN
ID ALU2_HUMAN STANDARD; PRT; 587 AA.
AC P39189;
DT 01-FEB-1995 (Rel. 31, Created)
DT 01-FEB-1995 (Rel. 31, Last sequence update)
DT 10-MAY-2005 (Rel. 47, Last annotation update)
DE Alu subfamily SB sequence contamination warning entry.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini; Hominidae;
OC Homo.
OX NCBI_TaxID=9606;
RN [1]
RP NUCLEOTIDE SEQUENCE.
RX MEDLINE=92241891; PubMed=1572661;
RA Claverie J.-M.;
RT "Identifying coding exons by similarity search: alu-derived and other
RL potentially misleading protein sequences.";
RN [3]
RP ALU FAMILIES CLASSIFICATION.
```

```
RX MEDLINE=88333009; PubMed=3138422;
RA Quent in Y.;
RT "The Alu family developed through successive waves of fixation closely
RL connected with primate lineage history.";
RN [4]
RP ALU FAMILIES CLASSIFICATION.
RX MEDLINE=91178815; PubMed=1706781;
RA Jurka J., Milosavljevic A.;
RT "Reconstruction and analysis of human Alu genes.";
RN [4]
RP ALU FAMILIES CLASSIFICATION.
CC -!- MISCELLANEOUS: Various analyses indicate that Alu repeats fall
CC into 8 subfamilies (Ref.3 and Ref.4). Therefore, 8 Alu warning
CC consensus sequences have been constituted that contain all six
CC frames conceptual translations of each of these classes of Alu
CC repeats.
CC -!- MISCELLANEOUS: Isolated 'X' indicates the presence of a stop
CC codon, 'XXX' is used to separate the various translation phases.
CC -!- CAUTION: This Alu entry is provided in order to avoid the further
CC pollution of protein sequence databases with Alu-derived amino
CC acid sequences.
CC -!- CAUTION: Alu repetitive sequences are interspersed in human and
CC primate genomes with an average spacing of 4 kb. Some of them are
CC actively transcribed by pol III. Normal transcripts may contain
CC Alu-derived sequences in 5' or 3' untranslated regions. However,
CC cDNA libraries also contain partial and/or rearranged cDNAs
CC ligated with Alu-derived sequence in any orientation. Although Alu
CC elements (especially situated on the complementary strand) have a
CC great potential to create additional/alternative exons,
CC consideration should be given to the possibility that the presence
CC of an Alu in an open reading frame may have resulted from a
CC cloning artifact or may be due to misinterpretation of sequencing
CC data. This point has been overlooked on several occasions, with
CC the consequence of erroneous Alu-derived amino acid sequences
CC being reported.
CC -!- CAUTION: Any significant similarity of a putative protein sequence
CC with an Alu-translated entry must be taken as a warning that a
CC part of Alu repeat may have been artifactually included in the
CC coding nucleotide sequence.
CC -----
CC This Swiss-Prot entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL Outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use as long as its content is in no way modified and this statement is not
CC removed.
CC -----
CC EMBL; U14568; -; NOT_ANNOTATED_CDS; Genomic_DNA.
CC REGION 1 96
CC REGION 100 194 Frame-1.
CC REGION 198 292 Frame-2.
CC REGION 296 391 Frame-3.
CC REGION 395 489 Frame-4.
CC REGION 493 587 Frame-5.
CC REGION 587 587 Frame-6.
CC SEQUENCE 587 AA; 63742 MW; 3EAB3E3E3929203 CRC64;

Alignment Scores:
Pred. No.: 7,51e-30 Length: 587
Score: 362.00 Matches: 92
Percent Similarity: 59.5% Conservative: 11
Best Local Similarity: 53.2% Mismatches: 41
Query Match: 40.2% Indels: 29
DB: 1 Gaps: 4

US-10-607-806-1_COPY_5000_5500 (1-501) x ALU2_HUMAN (1-587)
QY 499 AACCCCTCTTTTATAAAATACAAAAATAGCGGGCGTGTGGCAGCGCTTTAG 440
DB 134 AsnProValSerThrLysAsnThr--LysIleSerArgAlaTrpTrpArgAlaProValV 153
QY 439 TCCAGCTACTGGGAGGCTGAGCGCGGAGAGATGGCGTGAACCCAGCAGTGAGCTTGC 380
DB 153 alProAlaThrArgGluAlaGluAlaGlyGluTrpArgGluProGlyArgSerLeuG 173
```



```

Db 401 -----
QY 324 ACAGAGTCTTGTCTTGTTCGCCAAGCTGGAGTACAGTGTGCGATCTCGGCTCACTCGCA 383
Db 403 GlyValSer-LeuLeuLeuProArgLeuGluCysSerGlyAlaIleSerAlaHisCysAs 422
QY 384 GTCACACCTGCTGGGTTCACGCCATTCTCCGCCCTCAGCTCCGCCAGTAGCTGGACTAA 443
Db 422 nLeuArgLeuProGlySerSerAspSerProAlaSerAlaSerArgValAlaGlyIleTh 442
QY 444 AGCGCCTGCCACACCGCCCGCTAATTTTGTGTTTAAATAAAGACGGGTTT 500
Db 442 rGlyAlaArgHisAlaArgLeuIle--PheValPheLeuValGluThrGlyPhe 460

RESULT 7
Q6ZS15 HUMAN
ID Q6ZS15 HUMAN PRELIMINARY; PRT; 179 AA.
AC Q6ZS15;
DT 05-JUL-2004 (TREMBLrel. 27, Created)
DT 05-JUL-2004 (TREMBLrel. 27, Last sequence update)
DT 05-JUL-2004 (TREMBLrel. 27, Last annotation update)
DE Hypothetical protein FLJ45499.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini; Hominidae;
OC Homo.
OX NCBI_TaxID=9606;
RN [1]
RP NUCLEOTIDE SEQUENCE.
RC TISSUE=Thalamus;
RA Ota T., Nakagawa S., Senoh A., Mizuguchi H., Inagaki H., Sugiyama T.,
RA Irie R., Otsuki T., Sato H., Wakamatsu A., Ishii S., Yamamoto J.,
RA Isono Y., Kawai-Hio Y., Saito K., Nishikawa T., Kimura K.,
RA Yamashita H., Matsuo K., Nakamura Y., Sekine M., Kikuchi H., Kanda K.,
RA Wagatsuma M., Murakawa K., Kanehori K., Takahashi-Fujii A., Oshima A.,
RA Sugiyama A., Kurakami B., Suzuki Y., Sugano S., Nagahari K.,
RA Masuho Y., Nagai K., Isogai T.;
RL Submitted (JUL-2003) to the EMBL/GenBank/DBJ databases.
DR EMBL; AK127408; BAC86963.1; -; mRNA.
SQ SEQUENCE 179 AA; 19118 MW; 52286CCCA5A4450E CRC64;

Alignment Scores:
Pred. No.: 1.37e-29 Length: 179
Score: 358.50 Matches: 92
Percent Similarity: 61.3% Conservative: 11
Best Local Similarity: 54.8% Mismatches: 43
Query Match: 39.8% Indels: 22
DB: 2 Gaps: 5

US-10-607-806-1_COPY_5000_5500 (1-501) x Q6ZS15_HUMAN (1-179)
QY 12 TCTCTGCCAGGCTGGAGTCGGTGGTGCATCTCAGATCACTACACCTCGATCTCCT 71
Db 5 SerValSerGlnAlaGlyValGlnTrpLeuAspLeuSerSerLeuGlnProProSer 24
QY 72 GGGTTCAATAATCTCTTGCTCAGCTCTCAAGTAGCTGGAGTGTAGCTCTCAAGT 131
Db 25 GlyPheLysArgPheSerCysLeuSerLeuGlnSerTrpAspTyr----- 40
QY 132 AGCTGGCACACACACCATGCCAGCTAAATTTTGTGTGTTTTTGGTAGACACAGG 191
Db 41 --ArgHisAlaProProCysProAlaGly-PheCysLeu-----CysValGluThrGl 57
QY 192 TTTTACCATTGGCCGGCTGGTGACCTCCCTTTTAGATTCTCTATCTGCTCTA 251
Db 57 yPheHisValGlyGlnAlaSer-----LeuLysLeuLeuThrSe 71
QY 252 TTCTTCCCTTCTTAATGCAGTATCCAGTTTCCTTACTATCA-----CATTTATT 302
Db 71 rSerAspProProAlaLeuAlaSerGlnIleAlaGlyIleThrGlyMetIleHisAl 91
QY 303 ATTATCTTATTATTATTGACACAGAGTCTTGTCTTTCGCCAAGGCTGGAGTACAGTGG 362

```

```

Db 91 aGlnProLeuLeuPheLeuArgGlnGlyValThrLeuLeuProArgLeuGluCysSerSe 111
QY 363 TGGCATCTCGCTCAGCTGCAAGCTCCACCTGCTGGGTTCAGCCATTTCTCCGCTCAGC 422
Db 111 rAlaIleLeuAlaHisCysIleLeuCysLeuProGlySerSerAspSerProAlaSerAl 131
QY 423 CTCCTCCAGTAGCTGGGACTAAAGCGCTGTCACACACCGCCCGCTAAATTTTGTATT 482
Db 131 aSerGlnValAlaGlyThrThrGlyValCys-----ProLeuIle--PheValPhe 147
QY 483 TTAATAAAGACGGGTTT 500
Db 148 ValValGluMetGlyPhe 153

RESULT 8
ALU8_HUMAN
ID ALU8_HUMAN STANDARD; PRT; 591 AA.
AC P39195;
DT 01-FEB-1995 (Rel. 31, Created)
DT 01-FEB-1995 (Rel. 31, Last sequence update)
DT 10-MAY-2005 (Rel. 47, Last annotation update)
DE Alu subfamily SX sequence contamination warning entry.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini; Hominidae;
OC Homo.
OX NCBI_TaxID=9606;
RN [1]
RP NUCLEOTIDE SEQUENCE.
RX MEDLINE=9501759; PubMed=7935834; DOI=10.1038/371752a0;
RA Claverie J.-M., Makalowski W.;
RL "Alu alert.";
RN [2]
RX CONCEPT.
RX MEDLINE=92241891; PubMed=1572661;
RA Claverie J.-M.;
RT "Identifying coding exons by similarity search: alu-derived and other
RT potentially misleading protein sequences.";
RL Genomics 12:838-841(1992).
RN [3]
RX ALU FAMILIES CLASSIFICATION.
RX MEDLINE=88333009; PubMed=3138422;
RA Quentin Y.;
RT "The Alu family developed through successive waves of fixation closely
RT connected with primate lineage history.";
RL J. Mol. Evol. 27:194-202(1988).
RN [4]
RX ALU FAMILIES CLASSIFICATION.
RX MEDLINE=91178815; PubMed=1706781;
RA Jurka J., Milosavljevic A.;
RT "Reconstruction and analysis of human Alu genes.";
RL J. Mol. Evol. 32:105-121(1991).
CC -I- MISCELLANEOUS: Various analyses indicate that Alu repeats fall
CC into 8 subfamilies (Ref.3 and Ref.4). Therefore, 8 Alu warning
CC consensus sequences have been constituted that contain all six
CC frames conceptual translations of each of these classes of Alu
CC repeats.
CC -I- MISCELLANEOUS: Isolated 'X' indicates the presence of a stop
CC codon, 'XXY' is used to separate the various translation phases.
CC -I- CAUTION: This Alu entry is provided in order to avoid the further
CC pollution of protein sequence databases with Alu-derived amino
CC acid sequences.
CC -I- CAUTION: Alu repetitive sequences are interspersed in human and
CC primate genomes with an average spacing of 4 kb. Some of them are
CC actively transcribed by pol III. Normal transcripts may contain
CC Alu-derived sequences in 5' or 3' untranslated regions. However,
CC cDNA libraries also contain partial and/or rearranged cDNAs
CC ligated with Alu-derived sequence in any orientation. Although Alu
CC elements (especially situated on the complementary strand) have a
CC great potential to create additional/alternative exons,
CC consideration should be given to the possibility that the presence

```


of an Alu in an open reading frame may have resulted from a cloning artifact or may be due to misinterpretation of sequencing data. This point has been overlooked on several occasions, with the consequence of erroneous Alu-derived amino acid sequences being reported.

CAUTION: Any significant similarity of a putative protein sequence with an Alu-translated entry must be taken as a warning that a part of Alu repeat may have been artifactually included in the coding nucleotide sequence.

This Swiss-Prot entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use as long as its content is in no way modified and this statement is not removed.

EMBL: U14574; -; NOT_ANNOTATED_CDS; Genomic_DNA.

Hypothetical protein.

REGION 1	96
Frame-1.	
FT REGION 100	195
Frame-2.	
FT REGION 199	294
Frame-3.	
FT REGION 298	393
Frame-4.	
FT REGION 397	492
Frame-5.	
FT REGION 496	591
Frame-6.	
FT REGION 591	644
SEQUENCE	591 AA; 64434 MW; AC9154AD8A6B280 CRC64;

Alignment Scores:

Pred. No.:	1,03e-28	Length:	591
Score:	351.50	Matches:	94
Percent Similarity:	62.5%	Conservative:	11
Best Local Similarity:	56.0%	Mismatches:	45
Query Match:	39.1%	Indels:	18
DB:	1	Gaps:	6

US-10-607-806-1_COPY_5000_5500 (1-501) x ALU8_HUMAN (1-591)

QY 499 AACCCGCTCTTTATTAAATAACAAAAATAGCGGGCGTGGTGGCGAGCGCTTTAG 440

Db 36 AnProValSerThrLysAsnThr--LysIleSerArgAlaTrpArgAlaProValI 55

QY 439 TCCAGCTACTGGGAGGTGGCGGAGATGGCGTGAACCCAGCGTGGAGCTTGC 380

Db 55 leProAlaThrArgGluAlaGluAlaGlySerLeuGluProGlyArgArgGleu 75

QY 379 AGTGAGCCGAGATCGCACCTGTACTCCAGCTTGGCGACAAAGCAAGACTCTGTCTCA 320

Db 75 In***AlaGluIleAlaProLeuHisSerSerLeuGlyAspArgAlaArgLeu- 94

QY 319 ATAATAAGTAATAATAATAATGTGATAAGTAAGGAACCTGG-----ATACTGC 269

Db 95 -----LysLys*****AlaGlyGlyArgAlaArgTrpLeuThrProValIleP 113

QY 268 ATTGAAAGGGGAGAGATAGACGAGATAGGAGATCTAAAGGGAGGTCCACCGCCT 209

Db 113 roAlaLeuTrpGluGlu---AspHisLeuArg-SerGlyValArgasp-----GlnPro 129

QY 208 GGCACATGCTGCAAAACCTGTCTTACCAAAAAACACACAAAAATTAGTGGGCAT 149

Db 130 GlyGlnHisGlyGluThrProSerLeuLeuLys-----IleGlnLysLeuAlaGlyArg 147

QY 148 GGTGGTGTGGCCAGCTACTTGAGAGCTACAAGTCCAGCTACTTGAGAGCGTGGAGGAA 89

Db 148 GlyGlyAla-----ArgLeu***SerGlnLeuLeuGlyArgLeuArgGln 162

QY 88 GAGAAATTATTGAACCCAGAGATGAGGTGTAGTGTAGTCTGAGATCGACCGCCACT 29

Db 163 GluAsnArgLeuAsnProGlyGlyGlyCysSerGluProArgSerArgHisCysThr 182

QY 28 CCAGCTGGCGCACAGAG 11

Db 183 ProAlaTrpAlaThrGlu 188

ALU3_HUMAN

ID ALU3_HUMAN STANDARD; PRT; 587 AA.

AC P39190;

DT 01-FEB-1995 (Rel. 31, Created)

DT 01-FEB-1995 (Rel. 31, Last sequence update)

DT 10-MAY-2005 (Rel. 47, Last annotation update)

DE Alu subfamily SBI sequence contamination warning entry.

OS Homo sapiens (Human)

OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini; Hominidae; Homo.

OC NCBI_TaxID=9606;

OX [1]

RN NUCLEOTIDE SEQUENCE.

RP MEDLINE=95021758; PubMed=7935834; DOI=10.1038/371752a0;

RA Claverie J.-M.; Makalowski W.;

RT "Alu alert.";

RL Nature 371:752-752 (1994).

RN [2]

RP CONCEPT.

RP MEDLINE=92241891; PubMed=1572661;

RX Claverie J.-M.;

RA "Identifying coding exons by similarity search: alu-derived and other potentially misleading protein sequences.";

RT Genomics 12:838-841 (1992).

RL [3]

RN ALU FAMILIES CLASSIFICATION.

RP MEDLINE=88333009; PubMed=3138422;

RA Quentin Y.;

RT "The Alu family developed through successive waves of fixation closely connected with primate lineage history.";

RL J. Mol. Evol. 27:194-202 (1988).

RN [4]

RP ALU FAMILIES CLASSIFICATION.

RP MEDLINE=91178815; PubMed=1706781;

RA Jurka J.; Milosavljevic A.;

RT "Reconstruction and analysis of human Alu genes.";

RL J. Mol. Evol. 32:105-121 (1991).

CC -!- MISCELLANEOUS: Various analyses indicate that Alu repeats fall into 8 subfamilies (Ref.3 and Ref.4). Therefore, 8 Alu warning consensus sequences have been constituted that contain all six frames conceptual translations of each of these classes of Alu repeats.

CC -!- MISCELLANEOUS: Isolated 'X' indicates the presence of a stop codon, 'XXX' is used to separate the various translation phases.

CC -!- CAUTION: This Alu entry is provided in order to avoid the further pollution of protein sequence databases with Alu-derived amino acid sequences.

CC -!- CAUTION: Alu repetitive sequences are interspersed in human and primate genomes with an average spacing of 4 kb. Some of them are actively transcribed by pol III. Normal transcripts may contain Alu-derived sequences in 5' or 3' untranslated regions. However, cDNA libraries also contain partial and/or rearranged cDNAs ligated with Alu-derived sequence in any orientation. Although Alu elements (especially situated on the complementary strand) have a great potential to create additional/alternative exons, consideration should be given to the possibility that the presence of an Alu in an open reading frame may have resulted from a cloning artifact or may be due to misinterpretation of sequencing data. This point has been overlooked on several occasions, with the consequence of erroneous Alu-derived amino acid sequences being reported.

CC -!- CAUTION: Any significant similarity of a putative protein sequence with an Alu-translated entry must be taken as a warning that a part of Alu repeat may have been artifactually included in the coding nucleotide sequence.

CC -----

CC This Swiss-Prot entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use as long as its content is in no way modified and this statement is not removed.

CC -----


```
DR EMBL; U14569; -, NOT_ANNOTATED_CDS; Genomic_DNA.
KW Hypothetical protein.
FT REGION 1 96 Frame-1.
FT REGION 100 194 Frame-2.
FT REGION 198 292 Frame-3.
FT REGION 296 391 Frame-4.
FT REGION 395 489 Frame-5.
FT REGION 493 587 Frame-6.
SQ SEQUENCE 587 AA; 63613 MW; 85C4155726DEF235 CRC64;

Alignment Scores:
Pred. No.: 1,49e-28 Length: 587
Score: 350.00 Matches: 102
Percent Similarity: 43.7% Conservatives: 12
Best Local Similarity: 39.1% Mismatches: 43
Query Match: 39.9% Indels: 105
DB: 1 Gaps: 8

US-10-607-806-1_COPY_5000_5500 (1-501) x ALU3_HUMAN (1-587)
QY 7 CTCACCTCTGTCGCCAGGCTGGAGTGGCGTGGATCTCAGATCACTACAACCTCCAT 66
Db 302 LeuAlaLeuSerProArgLeuGluCysSerGlyGlyIleSerAlaHisCysIleLeuArg 321
QY 67 CTCCTGGGTCAAAATATCTCTGCTCAGCTCTCAGTAGCTGGAGCTGTAGCTCT 126
Db 322 LeuProGlySerArgHisSerProAlaSerAlaSerAlaSerGlnValAlaGlyThr 338
QY 127 CAAGTAGCTGGCACACACACCATGCCAGCTAAATTTTGTGTTTTTGTGAGAG 186
Db 339 -----ThrGlyAlaArgHisTyrAlaArgLeuIlePhe-----ValPheLeuValGlu 354
QY 187 ACAGGTTTTTCACCA----- 200
Db 355 ThrGlyPheHisArgPheSerArgAspGlyLeuAspLeuLeuThrSer***SerAlaArg 374
QY 201 -----TGTTGG----- 206
Db 375 LeuGlyLeuProLysCysTyrAspTyrArgArgGluProProArgProAla***** 394
QY 207 -----CCAGGCTGG----- 215
Db 395 PhePhe***AspGlyValSerLeuCysArgProGlyTyrSerAlaValAlaGlySerArg 414
QY 216 GTGACCTCC-----CTTTTAGATTCTCTCT----- 239
Db 415 LeuThrAlaSerSerAlaSerArgValHisAlaIleLeuLeuProGlnProLys*** 434
QY 240 -----CATCTGCTCTATTCTTCTCCCTTTCTAATGAGTAT 275
Db 435 LeuGlyLeuGlnAlaProAlaThrThrProGly***PheLeuTyrPhe*****ArgArg 454
QY 276 CCAGTTTCTTACTTATACATTTATTATTATT----- 308
Db 455 GlyPheThrValLeuAlaGlyMetValSerIleSer***ProArgAspProProAlaSer 474
QY 309 -----CTTATTATTATT 320
Db 475 AlaSerGlnSerAlaGlyIleThrGlyValSerHisArgAlaArg*****PhePhe 494
QY 321 GAGACAGAGTCTTGCTTTGTCGCCAAGGCTGGAGTACAGTGGTGGATCTCGGCTCACTG 380
Db 495 GluThrGluSerArgSerValAlaGlnAlaGlyValGlnTyrArgAspLeuGlySerLeu 514
QY 381 CAAGCTCACTGTGGGTTACGCCATTCTCCGCTTCAGCTCCCGCTCCCGAGTGGGAC 440
Db 515 GlnAlaProProProGlyPheThrProPheSerCysLeuSerLeuProSerSerTyrAsp 534
QY 441 TAAAGGCGCTGCCACACCGCCGCTAAATTTTGTGTTTTTGTGTTTTTAAAGCGGCTT 499
Db 535 TyrArgArgProProLeuArgProAlaAsn---PheCysIlePheSerArgAspGlyVal 553

RESULT 10
```

QB9NAL9_HUMAN

```
ID QB9NAL9_HUMAN PRELIMINARY; PRT; 158 AA.
AC QB9NAL9;
DT 01-OCT-2002 (TrEMBLrel. 22, Created)
DT 01-OCT-2002 (TrEMBLrel. 22, Last sequence update)
DT 01-OCT-2002 (TrEMBLrel. 22, Last annotation update)
DE Hypothetical protein FLJ35131.
OS Homo sapiens (Human)
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini; Hominoidea;
OC Homo
OC NCBI_TaxID=9606;
OX [1]
RN NUCLEOTIDE SEQUENCE.
RC TISSUE=Placenta;
RX PubMed=14702019; DOI=10.1038/ng1285;
RA Ota T., Suzuki Y., Nishikawa T., Otsuki T., Sugiyama T., Irie R.,
RA Wakamatsu A., Hayashi K., Sato H., Nagai K., Kimura K., Makita H.,
RA Sekine M., Obayashi M., Nishi T., Shibahara T., Tanaka T., Ishii S.,
RA Yamamoto J.-I., Saito K., Kawai Y., Isono Y., Nakamura Y.,
RA Nagahari K., Murakami K., Yasuda T., Iwayanagi T., Wagatsuma M.,
RA Shiratori A., Sudo H., Hosoiri T., Kaku Y., Kodaira H., Kondo H.,
RA Sugawara M., Takahashi M., Kanda K., Yokoi T., Furuya T., Kikkawa E.,
RA Omura Y., Abe K., Kamihara K., Katsuta N., Sato K., Tanikawa M.,
RA Yamazaki M., Ninomiya K., Ishibashi T., Yamashita H., Murakawa K.,
RA Fujimori K., Tanai H., Kimata M., Watanabe M., Hiraoaka S., Chiba Y.,
RA Ishida S., Ono Y., Takiguchi S., Watanabe S., Yosida M., Hotuta T.,
RA Kusano J., Kanehori K., Takahashi-Fujii A., Hara H., Tanase T.-O.,
RA Nomura Y., Togishi S., Komai F., Hara R., Takeuchi K., Arita M.,
RA Imose N., Musashino K., Yuuki H., Oshima A., Sasaki N., Aotsuka S.,
RA Yoshikawa Y., Matsunawa H., Ichihara T., Shiohata N., Sano S.,
RA Moriya S., Momiyama H., Satoh N., Takami S., Terashima Y., Suzuki O.,
RA Nakagawa S., Senoh A., Mizoguchi H., Goto Y., Shimizu F., Wakebe H.,
RA Hishigaki H., Watanabe T., Sugiyama A., Takemoto M., Kawakami B.,
RA Yamazaki M., Watanabe K., Kumagai A., Itakura S., Fukuzumi Y.,
RA Fujimori Y., Komiyama M., Tashiro H., Tanigami A., Fujiwara T.,
RA Ono T., Yamada K., Fujii Y., Ozaki K., Hixao M., Ohmori Y.,
RA Kawabata A., Hikiji T., Kobatake N., Inagaki H., Ikema Y., Okamoto S.,
RA Okitani R., Kawakami T., Noguchi S., Itoh T., Shigeta K., Senba T.,
RA Matsumura K., Nakajima Y., Mizuno T., Morinaga M., Sasaki M.,
RA Togashi T., Oyama M., Hata H., Watanabe M., Komatsu T.,
RA Mizushima-Sugano J., Satoh T., Shirai Y., Takahashi Y., Nakagawa K.,
RA Okumura K., Nagase T., Nomura N., Kikuchi H., Masuho Y., Yamashita R.,
RA Nakai K., Yada T., Nakamura Y., Ohara O., Isoqai T., Sugano S.;
RT "Complete sequencing and characterization of 21,243 full-length human
RT cDNAs";
RL Nat. Genet. 36:40-45(2004).
DR EMBL; AK092450; BAC03893.1; -, mRNA.
DR Ensembl; ENSG00000198057; Homo sapiens.
SQ SEQUENCE 158 AA; 17414 MW; E151503015P2FE34 CRC64;
```

Alignment Scores:

```
Pred. No.: 4,36e-28 Length: 158
Score: 344.50 Matches: 80
Percent Similarity: 53.3% Conservatives: 8
Best Local Similarity: 48.5% Mismatches: 28
Query Match: 38.3% Indels: 49
DB: 2 Gaps: 4
```

US-10-607-806-1_COPY_5000_5500 (1-501) x QB9NAL9_HUMAN (1-158)

```
QY 6 GCTCACTCTGTCGCCAGGCTGGAGTGGCGTGGATCTCAGATCACTACAACCTCCA 65
Db 3 SerHisSerValThrGlnAlaGlyValGlnTyrPheGlyLeuGlySerLeuGlnProLeu 22
QY 66 TCTCTCTGGGTCAAAATATCTCTGCTCAGCTCTCAGTAGCTGGAGCTTTGTAGCTC 125
Db 23 ProProGlyPheIleArgPhePheCysLeuSerLeuLeuSerSerTyrPheTyr----- 40
QY 126 TCAAGTAGCTGGCACACACACCATGCCAGCTAAATTTTGTGTTTTTGTGTTGTTAGTA 185
Db 41 -----ArgArgValProProProAlaAsnPhe---CysIlePhePhe----- 54
```

```
QY 186 GACAGTTTTCACCAATGTTGCCAGGCTGGGTGACCTCCCTTTAGATTCTCCTCATCCT 245
Db 54 -----
QY 246 GCTCTATTCTTCCCTTCTTAATGAGATCCAGTTCTCCTTATCATTATTTATT 305
Db 55 SerPhePhePhe-----
QY 306 ATTCTTATTATTATGACAGAGTCTTGTTCGCCAAGCTGGAGTACAGTGGTGC 365
Db 59 -----
QY 366 GATCTCGGCTACTCAGAGTCCACTCTGCTGGTTTCAGCCATCTCCGCTCAGCTC 425
Db 74 AspLeuGlySerLeuGlnAlaAalaProProGlyPheThrProPheSerCysLeuSerLeu 93
QY 426 CCCAGTACTGGAGTAAAGGCGCTGCCACACGCCCGCTAAATTTTGTATTTTTA 485
Db 94 ProSerSerTrpAsnTrpArgProProProCysProAlaAsnPheValPheLeu 113
QY 486 ATAAAGACGGGGTTT 500
Db 114 ValGluMetGlyPhe 118
```

RESULT 11

```
ALU1_HUMAN
ID ALU1_HUMAN STANDARD; PRT; 591 AA.
AC P3188;
DT 01-FEB-1995 (Rel. 31, Created)
DT 01-FEB-1995 (Rel. 31, Last sequence update)
DT 10-MAY-2005 (Rel. 47, Last annotation update)
DE Alu subfamily J sequence contamination warning entry.
OS Homo sapiens (Human)
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini; Hominidae;
OC Homo.
OX NCBI_TaxID=9606;
RN [1]
RP NUCLEOTIDE SEQUENCE.
RX MEDLINE=95021758; PubMed=7935834; DOI=10.1038/371752a0;
RA Claverie J.-M., Makalowski W.;
RT "Alu alert.";
RL Nature 371:752-752(1994).
RN [2]
RP CONCEPT.
RX MEDLINE=92241891; PubMed=1572661;
RA Claverie J.-M.,
RT "Identifying coding exons by similarity search: alu-derived and other
potentially misleading protein sequences.";
RL Genomics 12:838-841(1992).
RN [3]
RP ALU FAMILIES CLASSIFICATION.
RX MEDLINE=88333009; PubMed=3138422;
RA Quentin Y.;
RT "The Alu family developed through successive waves of fixation closely
connected with primate lineage history.";
RL J. Mol. Evol. 27:194-202(1988).
RN [4]
RP ALU FAMILIES CLASSIFICATION.
RX MEDLINE=91178815; PubMed=1706781;
RA Jurka J., Milosavljevic A.;
RT "Reconstruction and analysis of human Alu genes.";
RL J. Mol. Evol. 32:105-121(1991).
CC -1- MISCELLANEOUS: Various analyses indicate that Alu repeats fall
into 8 subfamilies (Ref.3 and Ref.4). Therefore, 8 Alu warning
consensus sequences have been constituted that contain all six
frames conceptual translations of each of these classes of Alu
repeats.
CC -1- MISCELLANEOUS: Isolated 'X' indicates the presence of a stop
codon, 'XXX' is used to separate the various translation phases.
CC -1- CAUTION: This Alu entry is provided in order to avoid the further
pollution of protein sequence databases with Alu-derived amino
acid sequences.
```

```
CC -1- CAUTION: Alu repetitive sequences are interspersed in human and
primate genomes with an average spacing of 4 kb. Some of them are
actively transcribed by pol III. Normal transcripts may contain
Alu-derived sequences in 5' or 3' untranslated regions. However,
cDNA libraries also contain partial and/or rearranged cDNAs
ligated with Alu-derived sequence in any orientation. Although Alu
elements (especially situated on the complementary strand) have a
great potential to create additional/alternative exons,
consideration should be given to the possibility that the presence
of an Alu in an open reading frame may have resulted from a
cloning artifact or may be due to misinterpretation of sequencing
data. This point has been overlooked on several occasions, with
the consequence of erroneous Alu-derived amino acid sequences
being reported.
CC -1- CAUTION: Any significant similarity of a putative protein sequence
with an Alu-translated entry must be taken as a warning that a
part of Alu repeat may have been artifactually included in the
coding nucleotide sequence.
```

```
CC This Swiss-Prot entry is copyright. It is produced through a collaboration
between the Swiss Institute of Bioinformatics and the EMBL outstation -
the European Bioinformatics Institute. There are no restrictions on its
use as long as its content is in no way modified and this statement is not
removed.
```

DR EMBL; U14567; -, NOT_ANNOTATED_CDS; Genomic_DNA.

```
KW Hypothetical protein.
FT REGION 1 96 Frame-1.
FT REGION 100 195 Frame-2.
FT REGION 199 294 Frame-3.
FT REGION 298 393 Frame-4.
FT REGION 397 492 Frame-5.
FT REGION 496 591 Frame-6.
SQ SEQUENCE 591 AA; 63827 MW; 665D395735519D95 CRC64;
```

Alignment Scores:

```
Pred. No.: 9.7e-28 Length: 591
Score: 342.50 Matches: 93
Percent Similarity: 63.5% Conservative: 15
Best Local Similarity: 54.7% Mismatches: 46
Query Match: 38.1% Indels: 16
DB: 1 Gaps: 5
```

US-10-607-806-1_COPY_5000_5500 (1-501) x ALU1_HUMAN (1-591)

```
QY 4 GGGTCACTCTGTGCGCCAGCTGGAGTGGCGGTGATCTCAGATCACTACACCTC 63
Db 303 GlyLeuAlaLeuSerProArgLeuGluCysSerGlyAlaIleThrAlaHisCysSerLeu 322
QY 64 CATCTCTGGGTTCAATAATCTCTTGGCTCAGCTCTCAAGTAGCTGGGACTTGTAGC 123
Db 323 AspLeuProGlySerSerProAlaSerAlaSerArgValAlaGlyThr----- 340
QY 124 TCTCAAGTAGCTGGGCACACACACCATGCCAGCTAAATTTTGTGTGTTTTTGGTA 183
Db 341 -----ThrGlyAlaArgHisHisAlaArgLeuIlePhe---ValPhePhe---Val 355
QY 184 GAGACAGGTTTTCCACCATGTTGGCCAGCTGGTGAACCTCCCTTTTAGATTCTCTCATC 243
Db 356 GluThrGlyPheHisTy-ValAlaGlnAlaGlyLeuGlu-----LeuLeu-Gl 371
QY 244 CTGCTCTATTCTTCCCTTTCTTAATGAGTATCCAGTTTCTCTTACT-----TATCAT 297
Db 371 YserSerAspProProAlaSerAlaSerGlnSerAlaGlyIleThrGlyValSerHisAr 391
QY 298 TTATTATTATTCTTATTATTATTGAGACAGAGTCTTGTCTTTCGCCAAGGCTGGAGTAC 357
Db 391 galaarg*****PhePhe***AspArgValSerLeuCysArgProGlyTrpSerAl 411
QY 358 AGTGTGGCATCTCGGCTACATGCAAGCTTCCACTGCTGGTTCACGCCATTCTCCGCC 417
Db 411 aValAlaArgSerArgLeuThrAlaAlaSerThrSerArgAlaGlnAlaIleLeuLeuPr 431
```


GenCore version 5.1.7
Copyright (c) 1993 - 2006 Bioacceleration Ltd.

OM nucleic - protein search, using frame_plus_n2p model

Run on: February 8, 2006, 09:18:47 ; Search time 16.3058 Seconds
(without alignments)
2694.617 Million cell updates/sec

Title: US-10-607-806-1_COPY_1_500

Perfect score: 928

Sequence: 1 gacctaactgcatttg.....accacataagaggactga 500

Scoring table:

BLOSUM62
Xgapop 10.0 , Xgapext 0.5
Ygapop 10.0 , Ygapext 0.5
Fgapop 6.0 , Fgapext 7.0
Delop 6.0 , Delext 7.0

Searched: 2443163 seqs, 439378781 residues

Total number of hits satisfying chosen parameters: 4886326

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Command line parameters:

-MODEL=frame+ n2p.model -DEV=xlp
-Q=/abs/ABSSWEB/spool/US10607806/runat_08022006_091817_26823/app_query.fasta_1
-DB=A_Geneseq -QFMT=fastan -SUFFIX=n2p.rag -MINMATCH=0.1 -LOOPEXT=0
-UNITS=bits -START=1 -END=1 -MATRIX=blosum62 -TRANS=human40.cdi -LIST=45
-DOCALIGN=200 -THR SCORE=pct -THR MAX=100 -THR MIN=0 -ALIGN=15 -MODE=LOCAL
-OUTFMT=ptio -NORM=ext -HEAPSIZE=500 -MINLEN=0 -MAXLEN=2000000000 -HOST=abs02p
-USER=US10607806 @CGN 1.1 605 @runat_08022006_091817_26823 -NCPU=6 -ICPU=3
-NO_MMAP -NEG SCORES=0 -WAIT -DSPBLOCK=100 -LONGLOG -DRV TIMEOUT=120
-WARN TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5 -FGAPOP=6 -FGAPEXT=7
-YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database :

- A_Geneseq 21:*
- 1: Geneseq1980s:*
- 2: Geneseq1990s:*
- 3: Geneseq2000s:*
- 4: Geneseq2001s:*
- 5: Geneseq2002s:*
- 6: Geneseq2003as:*
- 7: Geneseq2003bs:*
- 8: Geneseq2004s:*
- 9: Geneseq2005s:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
C 1	350	37.7	110	3 AAB43292	Aab43292 Human ORF
C 2	339.5	36.6	143	5 ADK36945	Adk36945 Novel hum
C 3	324	34.9	85	4 AAU20249	Aau20249 Human nov
C 4	321	34.6	119	4 AAM83429	Aam83429 Human imm
C 5	318	34.3	260	8 ADR09618	Adr09618 Human pro
C 6	317.5	34.2	132	4 AAM01209	Aam01209 Human pro
C 7	317.5	34.2	132	4 AAU69854	Aau69854 Human pro
C 8	317.5	34.2	132	4 ABU71745	Abu71745 Prostate
C 9	317.5	34.2	132	5 ABB95314	Abb95314 Human P77

C 10	317.5	34.2	132	6 ABR54426	ABR54426 Prostate
C 11	317.5	34.2	132	7 ADB14023	Adb14023 Human pro
C 12	317.5	34.2	132	7 ADG26439	Adg26439 Human pro
C 13	317.5	34.2	135	4 AAM01284	Aam01284 P789P spl
C 14	317.5	34.2	135	4 AAU69929	Aau69929 Human pro
C 15	317.5	34.2	135	4 ABU71820	Abu71820 Prostate
C 16	317.5	34.2	135	5 ABB95389	Abb95389 Human P77
C 17	317.5	34.2	135	6 ABR54501	ABR54501 Prostate
C 18	317.5	34.2	135	7 ADB14334	Adb14334 Human pro
C 19	317.5	34.2	135	7 ADG26750	Adg26750 Human pro
C 20	317	34.2	94	5 ABJ03693	Abj03693 Human ova
C 21	317	34.2	123	7 ADB63958	Adb63958 Human pro
C 22	317	34.2	136	6 ADA54195	Ada54195 Human pro
C 23	316	34.1	85	4 AAO11283	Aao11283 Human pol
C 24	314	33.8	158	7 ADB64893	Adb64893 Human pro
C 25	311.5	33.6	138	7 ADM05509	Adm05509 Human pro
C 26	311	33.5	102	4 AAO08307	Aao08307 Human pol
C 27	310	33.4	124	8 ADQ65229	Adq65229 Novel hum
C 28	309	33.3	103	4 AAO06180	Aao06180 Human pol
C 29	307.5	33.1	202	8 ADQ66597	Adq66597 Novel hum
C 30	307	33.1	103	4 AAU32477	Aau32477 Novel hum
C 31	305.5	32.9	122	8 ADQ65187	Adq65187 Novel hum
C 32	305	32.9	80	5 ABP60188	Abp60188 Human mac
C 33	304	32.9	118	7 ADM05260	Adm05260 Human pro
C 34	304	32.8	127	5 AAE20800	Aae20800 Human gen
C 35	304	32.8	127	5 ABG64678	Abg64678 Human alb
C 36	304	32.8	127	8 ADL77945	Adl77945 Albumin f
C 37	302.5	32.6	104	7 ADB65611	Adb65611 Human pro
C 38	302.5	32.6	527	7 ADD27133	Add27133 Human adi
C 39	301.5	32.5	166	8 ADQ66566	Adq66566 Novel hum
C 40	301	32.4	118	6 ADA54860	Ada54860 Human pro
C 41	300.5	32.4	217	7 ADC87151	Adc87151 Human GPC
C 42	299	32.2	81	4 AAU29944	Aau29944 Novel hum
C 43	298	32.1	102	7 ADM06071	Adm06071 Human pro
C 44	298	32.1	134	5 AAO21703	Aao21703 Human sec
C 45	298	32.1	136	7 ADB65293	Adb65293 Human pro

ALIGNMENTS

RESULT 1

AAB43292

ID AAB43292 standard; protein; 110 AA.

XX AC AAB43292;

XX AC AAB43292;

DT 08-FEB-2001 (first entry)

XX 08-FEB-2001 (first entry)

DE Human ORFX ORF3056 polypeptide sequence SEQ ID NO:6112.

Human; open reading frame; ORFX; detection; cytotstatic; hepatotropic;
vulnary; antipsoriatic; antiparkinsonian; nootropic; neuroprotective;
anticonvulsant; osteopathic; antiarthritis; immunosuppressant; cardiant;
immunostimulant; thrombolytic; coagulant; vasotropic; antidiabetic;
hypotensive; dermatological; immunosuppressive; antiinflammatory;
antiviral; antibacterial; antifungal; antirheumatic; antithyroid;
antinaemic; gene therapy; cancer; proliferative disorder; hypertension;
neurodegenerative disorder; osteoarthritis; graft vs host disease;
cardiovascular disease; diabetes mellitus; hypothyroidism; SCID; AIDS;
cholesterol ester storage; systemic lupus erythematosus; infection;
severe combined immunodeficiency; malaria; autoimmune disorder; asthma;
allergy; aplastic anaemia; nocturnal haemoglobinuria; burn; wound;
bone damage; cartilage damage; antinflammatory disease; coagulation;
thrombosis; contraceptive.

XX Homo sapiens.

OS Aab43292 Human ORF

XX Adk36945 Novel hum

XX Aau20249 Human nov

PN Aam83429 Human imm

XX Adr09618 Human pro

PD Aam01209 Human pro

XX Aau69854 Human pro

XX Abu71745 Prostate

XX Abb95314 Human P77

XX 31-MAR-2000; 2000WO-US008621.

PR	31-MAR-1999;	99US-0127607P.	
PR	02-APR-1999;	99US-0127636P.	
PR	05-APR-1999;	99US-0127728P.	
PR	30-MAR-2000;	2000US-00540763.	
XX			
PA	(CURA-) CURAGEN CORP.		
XX			
PI	Shimkets RA, Leach M;		
XX			
DR	WPI; 2000-602362/57.		
DR	N-PSDB; AAC77501.		
XX			
PT	Novel nucleic acids and peptides derived from open reading frame X,		
PT	useful for treating e.g. cancers, proliferative disorders,		
PT	neurodegenerative disorders and cardiovascular disease.		
XX			
PS	Claim 11; Page 5297-5298; 5507pp; English.		
XX			
CC	AAC74446 to AAC77606 encode the proteins given in AAB40237 to AAB43397,		
CC	which represent the human ORFX open reading frames 1 to 3161. The ORFX		
CC	sequences have activities such as: cytostatic; hepatotropic; vulnery;		
CC	antiprosiatric; antiparkinsonian; nootropic; neuroprotective; osteopathic;		
CC	anticonvulsant; antiarthritic; immunosuppressant; immunostimulant;		
CC	cardiant; thrombolytic; coagulant; vasotropic; antidiabetic; hypotensive;		
CC	dermatological; immunosuppressive; antiinflammatory; antibacterial;		
CC	antiviral; antifungal; antirheumatic; antithyroid; and antianaemic. The		
CC	sequences can be used for determining the presence of or predisposition		
CC	to, or preventing or treating pathological conditions associated with an		
CC	ORFX-associated disorder. The nucleic acids can be used to express ORFX		
CC	proteins in gene therapy vectors. The proteins and nucleic acids may be		
CC	used to treat cancers, proliferative disorders, neurodegenerative		
CC	disorders, osteoarthritis, graft vs host disease, cardiovascular disease,		
CC	diabetes mellitus, hypertension, hypothyroidism, cholesterol ester		
CC	storage, systemic lupus erythematosus, severe combined immunodeficiency		
CC	(SCID), AIDS, viral, bacterial or fungal infection, malaria, autoimmune		
CC	disorders, asthma, allergies, aplastic anaemia, burns, wounds, bone and		
CC	cartilage damage, nocturnal haemoglobinuria, antiinflammatory disease; to		
CC	enhance coagulation; to inhibit thrombosis; and as a contraceptive		
XX			
SQ	Sequence 110 AA;		
	Alignment Scores:		
	Pred. No.:	4.87e-33	Length: 110
	Score:	350.00	Matches: 63
	Percent Similarity:	73.6%	Conservative: 4
	Best Local Similarity:	69.2%	Mismatches: 24
	Query Match:	37.7%	Indels: 0
	DB:	3	Gaps: 0
	US-10-607-806-1_COPY_1_500 (1-500) x AAB43292 (1-110)		
QY	347 CTTTITTTCCAGCTGGAGTGGTGGTCAATCTTGGCTCACTGCAACCTCCACCTCC 288		
Db	16 LeuValAlaGlnAlaGlyVal***TrpHisSerLeuGlySerLeuGlnProLeuPro 35		
QY	287 AGGTTCAAGTGAATCTTCTGCCCCAGCCTCCTGTGTAGCTGGACTATAGGCAAGTGCCA 228		
Db	36 GlyPheLysGlnPheSerCysArgSerLeuProSerSerTrpAspTrpArgHisAlaPro 55		
QY	227 CCACACTGGCTAATTTTGTATTTTATTTATAGAGATAGTGTTTTGCATGTTGCCAGGC 168		
Db	56 ProArgGlnAlaAsnPheCysIlePheSerArgAspGlyValSerProCysTrpProGly 75		
QY	167 TGGTCTCAACTCCTGCTCTCAAGTGTATCCACCAACTCGGTCTCCCAAAGTCTGGGAA 108		
Db	76 TrpSerGlnTrpProAspLeuArgSerTrpHisLeuSerValProLysCysTrpAsp 95		
QY	107 TACAGACATGAGCCACCGCACCTGGCTGAATGAG 75		
Db	96 TyrArgArgGluProProHisLeuAlaTyrGlu 106		
	RESULT 2		
	ADK36945		

ID	ADK36945 standard; protein; 143 AA.
XX	
AC	ADK36945;
XX	
DT	06-MAY-2004 (first entry)
XX	
DE	Novel human polypeptide SeqID9027.
XX	
KW	antiarthritic; antiparkinsonian; neuroprotective; nootropic;
KW	immunosuppressive; cytostatic; antiprosiatric; antiinflammatory;
KW	antibacterial; antiviral; antifungal; antiparasitic; gene therapy;
KW	arthritis; Parkinson's; Alzheimer's; autoimmune disease; cancer;
KW	psoriasis; inflammatory bowel disease; infection; bacteria; virus;
KW	fungus; parasite; human.
XX	
OS	Homo sapiens.
XX	
Key	Location/Qualifiers
XX	
FT	Misc-difference 1. .143
FT	/label= OTHER
FT	/note= "OTHER= All Xaa's in this sequence are unknown
FT	amino acids or the site of a stop codon within the DNA
FT	sequence"
XX	
PN	WO200216439-A2.
XX	
PD	28-FEB-2002.
XX	
XX	05-MAR-2001; 2001WO-US004941.
XX	
PR	07-MAR-2000; 2000US-00519705.
PR	19-MAY-2000; 2000US-00574454.
XX	
PA	(HYSE-) HYSEQ INC.
XX	
PI	Tang YT, Liu C, Drmanac RT;
XX	
DR	WPI; 2002-280918/32.
XX	
PT	Isolated polynucleotide encoding bone marrow derived polypeptides useful
PT	for treating, e.g., Parkinson's, Alzheimer's, cancer, arthritis, Crohn's
PT	disease, and inflammatory bowel disease.
XX	
XX	Claim 20; SEQ ID NO 9027; 504pp; English.
PS	
XX	This invention relates to a novel isolated polynucleotide comprising a
CC	nucleotide sequence selected from one of 1680 sequences, a mature protein
CC	coding portion of them, an active domain of them and their complementary
CC	sequences. The invention may be useful for the production of compounds
CC	with an antiarthritic, antiparkinsonian, neuroprotective, nootropic,
CC	immunosuppressive, cytostatic, antiprosiatric, antiinflammatory, in
CC	antibacterial, antiviral, antifungal or antiparasitic activity. In
CC	addition, the disclosed sequences may be useful for gene therapy. The
CC	polypeptides or their antibodies are useful for treating many diseases
CC	such as arthritis, Parkinson's, Alzheimer's, autoimmune diseases, cancer,
CC	psoriasis, inflammatory bowel disease and infections caused by bacteria,
CC	viruses, fungi or parasites. The present sequence is that of a human
CC	polypeptide of the invention.
XX	
SQ	Sequence 143 AA;
	Alignment Scores:
	Pred. No.:
	Score:
	Percent Similarity:
	Best Local Similarity:
	Query Match:
	DB:
	US-10-607-806-1_COPY_1_500 (1-500) x ADK36945 (1-143)
QY	338 CAGGCTGGAGTGCAGTGGTGGCAATCTTGGCTCACTGCAACCTCCACCTCCAGGTTCAAG 279

Db 28 GlnGlyValGlnTrpCysAsnLeuSerSerLeuGlnProProLeuProGlyPheLys 47
QY 278 TGATTTCTTCCGCCAGCTCTCTGTAGTGGGACTATAGGCAAGTGCACACACCTG 219
Db 48 ***PheSerCysLeuSerLeuProSerSerTrpAspTyrArgCysValPro***HisPro 67
QY 218 GCTAATTTTGT-----ATTATTATAGAGATAGTGTGTTTGCCA 180
Db 68 ProAsnPheCysLeuPhePheIlePheTyrPheIlePheSerArgAspGluValSerPro 87
QY 179 TGTTCGCCAGGTGTCTCAAACTCTGCTCTCAAGTGATCCACCAACTCGGTCTCCA 120
Db 88 CysTrpProGlyTrpSerArgThrProAspLeuLysTrpSerThrArgLeuGlyLeuPro 107
QY 119 AAGTCTGGGATACAGACATGAGCCACCGCACCTGGCTAATGAGTTACTATTAACT 63
Db 108 LysCysTrpAspTyrArgHisGluProAlaAlaGlnProGluLysThrLeuLeuThr 126
RESULT 3
AAU20249
ID AAU20249 standard; protein; 85 AA.
AC AAU20249;
XX
DT 17-DEC-2001 (first entry)
XX
DE Human novel endocrine antigen, SEQ ID No 306.
XX
KW Human; endocrine antigen; cytostatic; antiinfertility; antidiabetic;
KW thyroid-active; adrenal-active; androgenic; gastric; gene therapy;
KW antisense-therapy; antibody; endocrine disorder; hormone imbalance;
KW reproductive disorder; endocrine cancer; pancreatic disorder;
KW diabetes mellitus; adrenal gland disorder; hirsutism; thyroid disorder;
KW hyperthyroidism; hypothalamic disorder; vanishing testes syndrome.
XX
OS Homo sapiens.
XX
XN WO200155319-A2.
XX
PD 02-AUG-2001.
XX
PF 17-JAN-2001; 2001WO-US0001335.
XX
XX 31-JAN-2000; 2000US-0179065P.
PR 04-FEB-2000; 2000US-0180628P.
PR 24-FEB-2000; 2000US-0184664P.
PR 02-MAR-2000; 2000US-0186350P.
PR 16-MAR-2000; 2000US-0189874P.
PR 17-MAR-2000; 2000US-0190076P.
PR 18-APR-2000; 2000US-0198123P.
PR 19-MAY-2000; 2000US-0205515P.
PR 07-JUN-2000; 2000US-0209467P.
PR 28-JUN-2000; 2000US-0214886P.
PR 30-JUN-2000; 2000US-0215135P.
PR 07-JUL-2000; 2000US-0216684P.
PR 07-JUL-2000; 2000US-0216880P.
PR 11-JUL-2000; 2000US-0217487P.
PR 11-JUL-2000; 2000US-0217496P.
PR 14-JUL-2000; 2000US-0218290P.
PR 26-JUL-2000; 2000US-0220963P.
PR 26-JUL-2000; 2000US-0220964P.
PR 14-AUG-2000; 2000US-0224518P.
PR 14-AUG-2000; 2000US-0224519P.
PR 14-AUG-2000; 2000US-0225213P.
PR 14-AUG-2000; 2000US-0225214P.
PR 14-AUG-2000; 2000US-0225266P.
PR 14-AUG-2000; 2000US-0225267P.
PR 14-AUG-2000; 2000US-0225268P.
PR 14-AUG-2000; 2000US-0225270P.
PR 14-AUG-2000; 2000US-0225447P.
PR 14-AUG-2000; 2000US-0225757P.
PR 14-AUG-2000; 2000US-0225758P.
PR 14-AUG-2000; 2000US-0225759P.
PR 18-AUG-2000; 2000US-0226279P.
PR 22-AUG-2000; 2000US-0226681P.
PR 22-AUG-2000; 2000US-0226686P.
PR 22-AUG-2000; 2000US-0227182P.
PR 23-AUG-2000; 2000US-0227009P.
PR 30-AUG-2000; 2000US-0228924P.
PR 01-SEP-2000; 2000US-0229287P.
PR 01-SEP-2000; 2000US-0229343P.
PR 01-SEP-2000; 2000US-0229344P.
PR 01-SEP-2000; 2000US-0229345P.
PR 05-SEP-2000; 2000US-0229509P.
PR 05-SEP-2000; 2000US-0229513P.
PR 06-SEP-2000; 2000US-0230437P.
PR 06-SEP-2000; 2000US-0230438P.
PR 08-SEP-2000; 2000US-0231242P.
PR 08-SEP-2000; 2000US-0231243P.
PR 08-SEP-2000; 2000US-0231244P.
PR 08-SEP-2000; 2000US-0231413P.
PR 08-SEP-2000; 2000US-0231414P.
PR 08-SEP-2000; 2000US-0232080P.
PR 12-SEP-2000; 2000US-0231968P.
PR 14-SEP-2000; 2000US-0232397P.
PR 14-SEP-2000; 2000US-0232398P.
PR 14-SEP-2000; 2000US-0232399P.
PR 14-SEP-2000; 2000US-0232400P.
PR 14-SEP-2000; 2000US-0232401P.
PR 14-SEP-2000; 2000US-0233063P.
PR 14-SEP-2000; 2000US-0233064P.
PR 14-SEP-2000; 2000US-0233065P.
PR 21-SEP-2000; 2000US-0234223P.
PR 21-SEP-2000; 2000US-0234274P.
PR 25-SEP-2000; 2000US-0234997P.
PR 25-SEP-2000; 2000US-0234998P.
PR 26-SEP-2000; 2000US-0235484P.
PR 27-SEP-2000; 2000US-0235834P.
PR 27-SEP-2000; 2000US-0235836P.
PR 29-SEP-2000; 2000US-0236327P.
PR 29-SEP-2000; 2000US-0236367P.
PR 29-SEP-2000; 2000US-0236368P.
PR 29-SEP-2000; 2000US-0236369P.
PR 29-SEP-2000; 2000US-0236370P.
PR 02-OCT-2000; 2000US-0236802P.
PR 02-OCT-2000; 2000US-0237037P.
PR 02-OCT-2000; 2000US-0237038P.
PR 02-OCT-2000; 2000US-0237039P.
PR 13-OCT-2000; 2000US-0239335P.
PR 13-OCT-2000; 2000US-0239337P.
PR 20-OCT-2000; 2000US-0240960P.
PR 20-OCT-2000; 2000US-0241221P.
PR 20-OCT-2000; 2000US-0241785P.
PR 20-OCT-2000; 2000US-0241786P.
PR 20-OCT-2000; 2000US-0241787P.
PR 20-OCT-2000; 2000US-0241808P.
PR 20-OCT-2000; 2000US-0241809P.
PR 01-NOV-2000; 2000US-0241826P.
PR 01-NOV-2000; 2000US-0244617P.
PR 08-NOV-2000; 2000US-0246474P.
PR 08-NOV-2000; 2000US-0246475P.
PR 08-NOV-2000; 2000US-0246476P.
PR 08-NOV-2000; 2000US-0246477P.
PR 08-NOV-2000; 2000US-0246478P.
PR 08-NOV-2000; 2000US-0246523P.
PR 08-NOV-2000; 2000US-0246524P.
PR 08-NOV-2000; 2000US-0246525P.
PR 08-NOV-2000; 2000US-0246526P.
PR 08-NOV-2000; 2000US-0246527P.
PR 08-NOV-2000; 2000US-0246528P.
PR 08-NOV-2000; 2000US-0246532P.
PR 08-NOV-2000; 2000US-0246609P.
PR 08-NOV-2000; 2000US-0246610P.
PR 08-NOV-2000; 2000US-0246611P.

PR 14-AUG-2000; 2000US-0225759P.
 PR 18-AUG-2000; 2000US-0226279P.
 PR 22-AUG-2000; 2000US-0226681P.
 PR 22-AUG-2000; 2000US-0226868P.
 PR 22-AUG-2000; 2000US-0227182P.
 PR 23-AUG-2000; 2000US-0227009P.
 PR 30-AUG-2000; 2000US-0228924P.
 PR 01-SEP-2000; 2000US-0229287P.
 PR 01-SEP-2000; 2000US-0229343P.
 PR 01-SEP-2000; 2000US-0229344P.
 PR 01-SEP-2000; 2000US-0229345P.
 PR 05-SEP-2000; 2000US-0229509P.
 PR 05-SEP-2000; 2000US-0229513P.
 PR 06-SEP-2000; 2000US-0230437P.
 PR 06-SEP-2000; 2000US-0230438P.
 PR 08-SEP-2000; 2000US-0231242P.
 PR 08-SEP-2000; 2000US-0231243P.
 PR 08-SEP-2000; 2000US-0231244P.
 PR 08-SEP-2000; 2000US-0231413P.
 PR 08-SEP-2000; 2000US-0231414P.
 PR 08-SEP-2000; 2000US-0232080P.
 PR 08-SEP-2000; 2000US-0232081P.
 PR 12-SEP-2000; 2000US-0231968P.
 PR 14-SEP-2000; 2000US-0232397P.
 PR 14-SEP-2000; 2000US-0232398P.
 PR 14-SEP-2000; 2000US-0232399P.
 PR 14-SEP-2000; 2000US-0232400P.
 PR 14-SEP-2000; 2000US-0232401P.
 PR 14-SEP-2000; 2000US-0233063P.
 PR 14-SEP-2000; 2000US-0233064P.
 PR 14-SEP-2000; 2000US-0233065P.
 PR 21-SEP-2000; 2000US-0234223P.
 PR 21-SEP-2000; 2000US-0234274P.
 PR 25-SEP-2000; 2000US-0234997P.
 PR 25-SEP-2000; 2000US-0234998P.
 PR 26-SEP-2000; 2000US-0234988P.
 PR 27-SEP-2000; 2000US-0235484P.
 PR 27-SEP-2000; 2000US-0235834P.
 PR 27-SEP-2000; 2000US-0235836P.
 PR 29-SEP-2000; 2000US-0236327P.
 PR 29-SEP-2000; 2000US-0236367P.
 PR 29-SEP-2000; 2000US-0236368P.
 PR 29-SEP-2000; 2000US-0236369P.
 PR 29-SEP-2000; 2000US-0236370P.
 PR 02-OCT-2000; 2000US-0236802P.
 PR 02-OCT-2000; 2000US-0237037P.
 PR 02-OCT-2000; 2000US-0237038P.
 PR 02-OCT-2000; 2000US-0237039P.
 PR 02-OCT-2000; 2000US-0237040P.
 PR 13-OCT-2000; 2000US-0239935P.
 PR 13-OCT-2000; 2000US-0239937P.
 PR 20-OCT-2000; 2000US-0240960P.
 PR 20-OCT-2000; 2000US-0241221P.
 PR 20-OCT-2000; 2000US-0241785P.
 PR 20-OCT-2000; 2000US-0241786P.
 PR 20-OCT-2000; 2000US-0241787P.
 PR 20-OCT-2000; 2000US-0241808P.
 PR 20-OCT-2000; 2000US-0241809P.
 PR 20-OCT-2000; 2000US-0241826P.
 PR 01-NOV-2000; 2000US-0244617P.
 PR 08-NOV-2000; 2000US-0246474P.
 PR 08-NOV-2000; 2000US-0246475P.
 PR 08-NOV-2000; 2000US-0246476P.
 PR 08-NOV-2000; 2000US-0246477P.
 PR 08-NOV-2000; 2000US-0246478P.
 PR 08-NOV-2000; 2000US-0246523P.
 PR 08-NOV-2000; 2000US-0246524P.
 PR 08-NOV-2000; 2000US-0246525P.
 PR 08-NOV-2000; 2000US-0246526P.
 PR 08-NOV-2000; 2000US-0246527P.
 PR 08-NOV-2000; 2000US-0246528P.
 PR 08-NOV-2000; 2000US-0246532P.
 PR 08-NOV-2000; 2000US-0246609P.
 PR 08-NOV-2000; 2000US-0246610P.

PR 08-NOV-2000; 2000US-0246611P.
 PR 08-NOV-2000; 2000US-0246613P.
 PR 17-NOV-2000; 2000US-0249207P.
 PR 17-NOV-2000; 2000US-0249208P.
 PR 17-NOV-2000; 2000US-0249209P.
 PR 17-NOV-2000; 2000US-0249210P.
 PR 17-NOV-2000; 2000US-0249211P.
 PR 17-NOV-2000; 2000US-0249212P.
 PR 17-NOV-2000; 2000US-0249213P.
 PR 17-NOV-2000; 2000US-0249214P.
 PR 17-NOV-2000; 2000US-0249215P.
 PR 17-NOV-2000; 2000US-0249216P.
 PR 17-NOV-2000; 2000US-0249217P.
 PR 17-NOV-2000; 2000US-0249218P.
 PR 17-NOV-2000; 2000US-0249244P.
 PR 17-NOV-2000; 2000US-0249245P.
 PR 17-NOV-2000; 2000US-0249264P.
 PR 17-NOV-2000; 2000US-0249265P.
 PR 17-NOV-2000; 2000US-0249297P.
 PR 17-NOV-2000; 2000US-0249299P.
 PR 01-DEC-2000; 2000US-0250160P.
 PR 01-DEC-2000; 2000US-0250391P.
 PR 05-DEC-2000; 2000US-0251030P.
 PR 05-DEC-2000; 2000US-0251988P.
 PR 05-DEC-2000; 2000US-0256719P.
 PR 08-DEC-2000; 2000US-0251479P.
 PR 08-DEC-2000; 2000US-0251856P.
 PR 08-DEC-2000; 2000US-0251868P.
 PR 08-DEC-2000; 2000US-0251869P.
 PR 08-DEC-2000; 2000US-0251989P.
 PR 11-DEC-2000; 2000US-0251990P.
 PR 05-JAN-2001; 2001US-0254057P.
 PR 05-JAN-2001; 2001US-0259678P.

(HUMA-) HUMAN GENOME SCI INC.

Rosen CA, Barash SC, Ruben SM;

WPI; 2001-483426/52.

N-PSDB; AAK56210.

Nucleic acids encoding human immune/hematopoietic antigen polypeptides, useful for preventing, diagnosing and/or treating cancers and metastasis.

Claim 11; SEQ ID NO 11022; 3071pp + Sequence Listing; English.

AAK54951 to AAK64702 encode the human immune/haematopoietic antigen (I) amino acid sequences given in AAK82170 to AAK91921. (I) have cytostatic activity, and can be used in gene therapy and vaccine production. (I) proteins and polynucleotides may be used in the prevention, diagnosis and treatment of diseases associated with inappropriate (I) expression. For example, they may be used to treat disorders associated with decreased expression by rectifying mutations or deletions in a patient's genome that affect the activity of (I) by expressing inactive proteins or to supplement the patient's own production of (I). Additionally, (I) polynucleotides may be used to produce the secreted (I), by inserting the nucleic acids into a host cell and culturing the cell to express the protein. (I) proteins and polynucleotides may be used to prevent, diagnose and treat immune/haematopoietic-related diseases, especially cancers and cancer metastases of haematopoietic-derived cells. AAK64703 to AAK87694 represent human immune/haematopoietic antigen genomic sequences from the present invention. AAK54942 to AAK54950 and AAK82169 represent sequences used in the exemplification of the present invention

SQ Sequence 119 AA;

Alignment Scores:

Pred. No.:	1,64e-29	Length:	119
Score:	321.00	Matches:	63
Percent Similarity:	74.2%	Conservative:	6
Best Local Similarity:	67.7%	Mismatches:	23
Query Match:	34.6%	Indels:	1

PI Xu J, Dillon DC, Mitcham JL, Harlocker SL, Jiang Y, Reed SG;
PI Kalos MD, Fanger GR, Day CH, Retter MW, Stolk JA, Skeiky YAW;
XX Wang A, Meagher MJ;
XX WPI; 2001-425873/45.
XX New polynucleotide encoding a prostate-specific protein, for diagnosing,
PT monitoring and treating prostate cancer in a patient and for use in
PT vaccines.
XX
XX
PS Claim 2; Page 435; 543pp; English.
XX
XX The present invention describes polynucleotide sequences (I) which encode
CC prostate-specific proteins (II). (I) and (II) have cytostatic activity,
CC and can be used in vaccine production and gene therapy. (I), (II),
CC antibodies to (II), fusion proteins comprising (II), and isolated T cells
CC prepared using (I) or (II) are used to treat cancer in a patient. (I) and
CC the antibodies are also used in the detection of cancer in a patient. The
CC cancer that is diagnosed or treated is particularly prostate cancer. (I)
CC and (II) can be used in vaccines. The antibodies or (I) can be used for
CC monitoring the progression of cancer in a patient. (I) and (II) can also
CC be used to improve diagnostic and therapeutic methods for prostate
CC cancer. They can indicate the level of metastasis as well as the prostate
CC volume. AAH93357 to AAH93944 and AAH0115 to AAH01318 represent
CC polynucleotide and amino acid sequences used in the exemplification of
CC the present invention
XX
SQ Sequence 132 AA;

Alignment Scores:
Pred. No.: 4,49e-29 Length: 132
Score: 317.50 Matches: 66
Percent Similarity: 61.1% Conservative: 14
Best Local Similarity: 50.4% Mismatches: 38
Query Match: 34.2% Indels: 13
DB: 4 Gaps: 2

US-10-607-806-1_COPY_1_500 (1-500) x AAH01209 (1-132)

QY 431 GGCTGAAGGATGCTAAGTGGGTGCTCAAGAGGAGGAGCAGCTATGG 372
Db 4 GlyGluGlyGluAlaArgHisValLeuHisGlyGlyArgGluArg----- 19
QY 371 TTGCAAGTAAAGTTATTAAATACCTCTTTTTC----- 339
Db 20 ValArgGlyGluThrAlaThrAsnPhePheLeuArgGlnGluSerGlyProValAla 39
QY 338 CAGCTGGAGTGCAGTGTGCAATCTTGCTGCTACCTGCACTCCACCTCCAGTTCAG 279
Db 40 GlnAlaGlyValGlnTrpHisAspLeuSerSerLeuGlnProLeuProHisArgPheLys 59
QY 278 TGATTCTTCTCCCGCAGCTCTGTGTAGCTGGGACTATAGGCAAGTGCACACACCTG 219
Db 60 GlnPheSerCysLeuSerLeuProHisSerTrpAspHisArgTyrAlaProProHisLeu 79
QY 218 GCTAATTTTGTATTTTATTAGATAGTGTGCTTTTGGCATGTGCGCAGGCTGCTCAA 159
Db 80 AlaAsnPheCysSerPheSerArgAspGlyValSerLeuCysCysSerGlyTrpSerLys 99
QY 158 ACTCCTGCTTCAAGTATGATCCACCACTCGGTCTCCCAAGTCTGGGAATACAGACAT 99
Db 100 ThrProGlyLeuGlnGlnSerAlaCysLeuGlyLeuProLysCysTrpGlyTyrArgHis 119
QY 98 GAGCCACGCGACCTGGCTAATGACTTACTATTA 66
Db 120 LysProProHisProAlaCysHisIleLeuLeu 130
RESULT 7
AAU69854
ID AAU69854 standard; protein; 132 AA.
XX
AC AAU69854;
XX

DT 30-JAN-2002 (first entry)
XX Human prostate cDNA encoded protein #49.
XX
XX Human; prostate cancer; cytostatic; immunostimulant; tumour; immunogen.
XX OS Homo sapiens.
XX WO200173032-A2.
XX
PD 04-OCT-2001.
XX
XX 27-MAR-2001; 2001WO-US009919.
XX 27-MAR-2000; 2000US-00536857.
XX 09-MAY-2000; 2000US-00568100.
XX 12-MAY-2000; 2000US-00570737.
XX 13-JUN-2000; 2000US-00593793.
XX 27-JUN-2000; 2000US-00605783.
XX 09-AUG-2000; 2000US-00636215.
XX 29-AUG-2000; 2000US-00651236.
XX 06-SEP-2000; 2000US-00657279.
XX 02-OCT-2000; 2000US-00679426.
XX 10-OCT-2000; 2000US-00685166.
XX 09-NOV-2000; 2000US-00709729.
XX
XX (CORI-) CORIXA CORP.
XX
XX Xu J, Dillon DC, Mitcham JL, Harlocker SL, Jiang Y, Kalos MD;
PI Fanger GR, Retter MW, Stolk JA, Day CH, Vedvick TS, Carter D;
PI Li SX, Wang A, Skeiky YAW, Hepler WT, Henderson RA;
XX
XX WPI; 2001-639232/73.
XX N-PSDB; AAS63925.
XX
XX New human prostate-specific polypeptides and polynucleotides useful for
PT the diagnosis and treatment of cancer, especially prostate cancer.
XX
XX Claim 2; Page 430-431; 579pp; English.
XX
XX The invention relates to isolated prostate-specific polynucleotides,
CC polypeptides, fusion proteins of the polypeptides, antibodies raised
CC against the polypeptides (or antigenic epitopes derived from them) and
CC antigen-presenting cells expressing the polypeptides. The antibodies are
CC useful for detecting the presence of cancer, especially prostate cancer.
CC The polypeptides, polynucleotides and the antigen-presenting cells are
CC useful for stimulating and/or expanding T cells specific for a tumour
CC protein, and for inhibiting the development of cancer especially prostate
CC cancer. Compositions comprising the polynucleotide and/or polypeptide are
CC useful for stimulating an immune response, and for treating cancer. The
CC oligonucleotide is useful for detecting cancer. The present sequence is a
CC prostate specific polypeptide of the invention
XX
SQ Sequence 132 AA;

Alignment Scores:
Pred. No.: 4,49e-29 Length: 132
Score: 317.50 Matches: 66
Percent Similarity: 61.1% Conservative: 14
Best Local Similarity: 50.4% Mismatches: 38
Query Match: 34.2% Indels: 13
DB: 4 Gaps: 2

US-10-607-806-1_COPY_1_500 (1-500) x AAU69854 (1-132)

QY 431 GGCTGAAGGATGCTAAGTGGGTGCTCAAGAGGAGGAGCAGCTATGG 372
Db 4 GlyGluGlyGluAlaArgHisValLeuHisGlyGlyArgGluArg----- 19
QY 371 TTGCAAGTAAAGTTATTAAATACCTCTTTTTC----- 339
Db 20 ValArgGlyGluThrAlaThrAsnPhePheLeuArgGlnGluSerGlyProValAla 39

XX	WPI:	2001-245062/25.
DR	N-PSDB;	ACA59733.
XX	Prostate specific protein and its encoding polynucleotide, useful for the treatment and diagnosis of prostate cancer.	
PT	Example 5;	SEQ ID NO 573; 85pp; English.
XX	The invention describes a fusion protein comprising at least one amino acid sequence of immunogenic portions of any of the 3 sequences not defined in the specification, or sequences having at least 70 or 90 % sequence identity to any one of the 35 sequences defined in the USPTO web site, which is encoded by any of the 4 nucleotide sequences not defined in the specification. The fusion protein, composition and methods are useful for diagnosing, preventing and/or treating cancer, particularly prostate cancer. The proteins are useful as markers to indicate the presence or absence of cancer. This is the amino acid sequence of a prostate cancer specific antigen. Note: The sequence data for this patent did not form part of the printed specification, but was obtained in electronic format directly from the US patent office at seqdata.uspto.gov/sequence.html?DocID=US20020192763	
XX	Sequence	132 AA;
SQ	Alignment Scores:	
	Pred. No.:	4.49e-29 Length: 132
	Score:	317.50 Matches: 66
	Percent Similarity:	61.1% Conservative: 14
	Best Local Similarity:	50.4% Mismatches: 38
	Query Match:	34.2% Indels: 13
DB:		Gaps: 2
US-10-607-806-1_COPY_1_500 (1-500) x ABU71745 (1-132)		
Qy	431	GGCCTGAAGGATGCTAAGTGGTGAATTGGGGTGGCTCAAAGAGGAGACGAGCTATGG 372
Db	4	GlyGluGlyGluAlaArgHisValLeuHisGlyGlyArgArgGluArg----- 19
Qy	371	TGCAAGTAAAGTTATTAAATACCTTTTTC----- 339
Db	20	ValArgGlyGluThrAlaThrAsnPhePheLeuArgGlnGluSerGlyProValAla 39
Qy	338	CAGCTGGAGTGCAGTGGTGCATCTTGGCTCACTGCAACCTCCACCTCCAGGTTCAAG 279
Db	40	GlnAlaGlyValGlnTrpHisAspLeuSerSerLeuGlnProLeuProHisArgPheLys 59
Qy	278	TGATTCTTCTGCCCGAGCTCCTGTGTAGCTGGGACTATAGGCAAGTGCACACACTG 219
Db	60	GlnPheSerCysLeuSerLeuProHisSerTrpAspHisArgTyrAlaProProHisLeu 79
Qy	218	GCTAATTTTGTATTTTATTAGAGATAGTGTGTTTGGCCATGTGGCCAGGCTGCTCAA 159
Db	80	AlaAsnPheCysSerPheSerArgAspGlyValSerLeuCysCysSerGlyTrpSerLys 99
Qy	158	ACTCTGCTCTCAAGTGAATCCCAACTCGGTCTCCCAAGTCTGCTGGGAATACAGACAT 99
Db	100	ThrProGlyLeuGlnGlnSerAlaCysLeuGlyLeuProLysCysTrpGlyTyrArgHis 119
Qy	98	GAGCCACCGCAGCTGGCTTAAGTACTATTA 66
Db	120	LysProProHisProAlaCysHisIleLeuLeu 130
RESULT 9		
ABB95314		
ID	ABB95314 standard; protein; 132 AA.	
XX	AC ABB95314;	
XX	DT 19-JUL-2002 (first entry)	
XX	DE Human P776P splice variant encoded protein SEQ ID NO 573.	
XX		

338	CAGGCTGGAGTGCAGTGGTGAATCTTGGCTCACTGCAACCTCCACCTCCAGGTTCAAG	279
40	GlnAlaGlyValGlnTrpHisAspLeuSerSerLeuGlnProLeuProHisArgPheLys	59
278	TGATTCTTCTGCCCGAGCTCCTGTGTAGCTGGGACTATAGGCAAGTGCACACACTG	219
60	GlnPheSerCysLeuSerLeuProHisSerTrpAspHisArgTyrAlaProProHisLeu	79
218	GCTAATTTTGTATTTTATTAGAGATAGTGTGTTTGGCCATGTGGCCAGGCTGCTCAA	159
80	AlaAsnPheCysSerPheSerArgAspGlyValSerLeuCysCysSerGlyTrpSerLys	99
158	ACTCTGCTCTCAAGTGAATCCCAACTCGGTCTCCCAAGTCTGCTGGGAATACAGACAT	99
100	ThrProGlyLeuGlnGlnSerAlaCysLeuGlyLeuProLysCysTrpGlyTyrArgHis	119
98	GAGCCACCGCAGCTGGCTTAAGTACTATTA 66	
120	LysProProHisProAlaCysHisIleLeuLeu 130	
SULT 8		
ABU71745		
ABU71745	standard; protein; 132 AA.	
ABU71745;		
10-JUN-2003	(first entry)	
Prostate cancer specific antigen P776P #1.		
Prostate cancer; vaccine; gene therapy; cytostatic; fusion protein;		
immunogen; cancer; prostate specific antigen.		
Homo sapiens.		
Synthetic.		
US2002192763-A1.		
19-DEC-2002.		
29-JUN-2001; 2001US-00895793.		
04-OCT-1999; 99US-0157455P.		
04-OCT-2000; 2000US-00679272.		
28-MAR-2001; 2001US-00822827.		
(XUJJ/) XU J.		
(DILL/) DILLON D C.		
(MITC/) MITCHAM J L.		
(HARL/) HARLOCKER S L.		
(JIAN/) JIANG Y.		
(KALO/) KALOS M D.		
(FANG/) FANGER G R.		
(RETT/) RETTER M W.		
(STOL/) STOLK J A.		
(DAYC/) DAY C H.		
(VEDV/) VEDVICK T S.		
(CART/) CARTER D.		
(LISX/) LI S X.		
(WANG/) WANG A.		
(SKEI/) SKEIKY Y A W.		
(HEPL/) HEPLER W T.		
(HEND/) HENDERSON R A.		
(HURA/) HURAL J.		
(MCNE/) MCNEILL P D.		
(HOUG/) HOUGHTON R L.		
(DBAS/) Y DE BASSOLS C V.		
(FOYT/) FOY T M.		
Xu J, Dillon DC, Mitcham JL, Harlocker SL, Jiang Y, Kalos MD;		
Fanger GR, Retter MW, Stolk JA, Day CH, Vedvick TS, Carter D;		
Li SX, Wang A, Skeiky YAW, Hepler WT, Henderson RA, Hural J;		
Mcneill PD, Houghton RL, Y De BassolsCV, Foy TW;		

PT prostate cancer in a patient, as well as for diagnosing prostate cancer
PT in a patient.
PS Example 5; Page 497-498; 691pp; English.
XX
XX The present invention relates to novel prostate-specific proteins (PSP)
CC and their coding sequences. The PSPs and their coding sequences are
CC useful for stimulating an immune response in a patient, or for treating
CC prostate cancer in a patient and for determining, detecting or diagnosing
CC the presence of a cancer in a patient. The present sequence was used to
CC illustrate the invention
XX
XX Sequence 132 AA;
SQ
Alignment Scores:
Pred. No.: 4,49e-29 Length: 132
Score: 317.50 Matches: 66
Percent Similarity: 61.1% Conservative: 14
Best Local Similarity: 50.4% Mismatches: 38
Query Match: 34.2% Indels: 13
DB: 6 Gaps: 2
US-10-607-806-1_COPY_1_500 (1-500) x ABR54426 (1-132)
QY 431 GGCCTGAAGGATCTAAGTGGTGAATGGGGTGGCTCAAGAGGAGGAGCAGCTATGG 372
Db 4 GlyGluGlyGluAlaArgHisValLeuHisGlyGlyArgArgGluArg----- 19
QY 371 TTGCAAGTAAAGTTTATTATACCTCTTTTTC----- 339
Db 20 ValArgGlyGluThrAlaThrAsnPhePheLeuArgGlnGluSerGlyProValAla 39
QY 338 CAGGCTGGAAGTGCAGTGGTGCATCTTGGCTCACTGCACCTCCACCTCCAGGTTCAAG 279
Db 40 GlnAlaGlyValGlnTrpHisAspLeuSerSerLeuGlnProLeuProHisArgPheLys 59
QY 278 TGATTCTTCTGCCAGCCTCTCTGTAGTGGGACTATAGCAAGTGCACACACACCTG 219
Db 60 GlnPheSerCysLeuSerLeuProHisSerTrpAspHisArgTyrAlaProProHisLeu 79
QY 218 GCTAATTTTGTATTATTATTAGATAGTGTGTTGCCATGTGCGCAGGCTGGTCTCAA 159
Db 80 AlaAsnPheCysSerPheSerArgAspGlyValSerLeuCysSerGlyTrpSerLys 99
QY 158 ACTCTGCTCTCAAGTGTATCCACCACTCGGTCTCCCAAGTGTCTGGGAATACAGACAT 99
Db 100 ThrProGlyLeuGlnGlnSerAlaCysLeuGlyLeuProLysCysTrpGlyTyrArgHis 119
QY 98 GAGCCACCGCCTGCTGCTAATGAGTTACTATTA 66
Db 120 LysProHisProAlaCysHisIleLeuLeu 130
RESULT 11
ADBI4023
ID ADBI4023 standard; protein; 132 AA.
AC
XX ADBI4023;
XX
DT 18-DEC-2003 (first entry)
XX
DE Human prostate specific cDNA P776P splice variant contig 6 protein #1.
XX
XX Human; prostate specific cDNA; cytostatic; immunostimulant; gene therapy;
KW cell therapy; vaccine; T-cell epitope;
KW class I major histocompatibility complex allele; MHC; prostate cancer;
KW tumour; antigen presenting cell.
XX
OS Homo sapiens.
XX
XX US2003185830-A1.
PN
XX
PD 02-OCT-2003.
XX

PF 12-NOV-2002; 2002US-00294025.
XX
PR 25-FEB-1997; 97US-00806099.
PR 01-AUG-1997; 97US-00904804.
PR 09-FEB-1998; 98US-00020956.
PR 25-FEB-1998; 98US-00030607.
PR 14-JUL-1998; 98US-00115453.
PR 23-SEP-1998; 98US-00159812.
PR 13-JAN-1999; 99US-00232149.
PR 09-APR-1999; 99US-00288946.
PR 13-JUL-1999; 99US-00352616.
PR 12-NOV-1999; 99US-00439313.
PR 18-NOV-1999; 99US-00443686.
PR 14-JAN-2000; 2000US-00483672.
PR 27-MAR-2000; 2000US-00536857.
PR 09-MAY-2000; 2000US-00568100.
PR 12-MAY-2000; 2000US-00570737.
PR 13-JUN-2000; 2000US-00593793.
PR 27-JUN-2000; 2000US-00605783.
PR 09-AUG-2000; 2000US-00636215.
PR 29-AUG-2000; 2000US-00651236.
PR 06-SEP-2000; 2000US-00657279.
PR 02-OCT-2000; 2000US-00679426.
PR 10-OCT-2000; 2000US-00685166.
PR 09-NOV-2000; 2000US-00709729.
PR 12-JAN-2001; 2001US-00759143.
PR 09-FEB-2001; 2001US-00780669.
PR 09-MAY-2001; 2001US-00852311.
PR 29-JUN-2001; 2001US-00895814.
PR 10-DEC-2001; 2001US-00012896.
PR 09-MAY-2002; 2002US-00144678.
XX
XX (CORI-) CORIXA CORP.
PA
XX Xu J, Stolk JA, Kalos MD;
PI
XX MPI; 2003-756193/71.
XX N-PSDB; ADBI4020.
DR
XX
XX New isolated polypeptide for use in a vaccine for stimulating an immune
PT response, or for treating or diagnosis cancer, preferably prostate
PT cancer.
XX
XX Example 5; Page; 101pp; English.
XX
XX The invention relates to an isolated polypeptide comprising no more than
CC 11-542 amino acids of ADBI3563 comprising a sequence ADBI4487. The
CC peptides comprise a fragment ADBI3563 of that contain naturally processed
CC T-cell epitopes for 3 class I major histocompatibility complex (MHC)
CC alleles. ADBI3563 is a polypeptide encoded by a human prostate specific
CC cDNA, one of 648 disclosed as new. Also included are nucleic acids
CC encoding the proteins and peptides, expression vectors, a host cell
CC transformed with the vector, an isolated antibody (or antigen binding
CC fragment) that specifically binds to the protein or peptide, detecting
CC the presence of a cancer in a patient (comprising contacting a patient
CC sample with a binding agent that binds to the peptides or a polypeptide
CC appearing as ADBI3558, detecting the amount of polypeptide that binds to
CC the agent and comparing the amount of polypeptide to a predetermined cut-
CC off value to determine the presence of cancer), a fusion protein
CC comprising the peptides or proteins, stimulating or expanding T cells
CC specific for a tumour protein comprising contacting T cells with the
CC peptides or the isolated T cell population, treating prostate cancer in a
CC patient comprising administering a composition comprising the presence of a
CC nucleic acids, antibodies or compounds, determining the presence of a
CC cancer in a patient and treating prostate cancer in a patient comprising
CC incubating cluster of differentiation (CD4+) and/or CD8+ T cells isolated
CC from a patient with the peptides or antigen presenting cells that express
CC (the peptides so that the T cells proliferate, and administering the
CC proliferated T cells to the patient. The peptides (or an oligonucleotide
CC that hybridises to nucleic acid encoding them), is used to detect the
CC presence of cancer in a patient. The peptides, nucleic acids encoding, or
CC antigen-presenting cells expressing the nucleic acid, are used to
CC stimulate or expand T cells specific for a tumour protein. The peptides,

FR	12-MAR-2000	2000US-005070373
FR	13-JUN-2000	2000US-005397393
FR	27-JUN-2000	2000US-005057833
FR	29-AUG-2000	2000US-006057813
FR	29-AUG-2000	2000US-006362153
FR	29-AUG-2000	2000US-006512363
FR	06-SEP-2000	2000US-0065727293
FR	02-OCT-2000	2000US-006794263
FR	10-NOV-2000	2000US-006851663
FR	09-NOV-2000	2000US-007097293
FR	13-JAN-2001	2001US-007591433
FR	09-FEB-2001	2001US-0078066593
FR	09-MAY-2001	2001US-008529113
FR	29-JUN-2001	2001US-008958143
FR	10-DEC-2001	2001US-009128963

PR 09-FEB-2001; 2001US-00780689.
PR 09-MAY-2001; 2001US-00852911.
PR 29-JUN-2001; 2001US-00895814.
PR 10-DEC-2001; 2001US-00012896.

PR 29-JUN-2001; 2001US-00895814.
PR 10-DEC-2001; 2001US-00012896.

XX
PA (CORI-) CORIXA CORP.

PI Xu J, Dillon DC, Mitcham JL

PI Kalos MD, Fanger GR, Retter
PI Carter D, Li SX, Wang A, S
PI Mancill MD, Houghton D, W

PI Meagher MJ, Deng T;
XX

DR WPT; 2003-111913/13.
DR N-PSDB; ADG26436.
XX

PT New polynucleotides encoding
PT a human prostate tumor cDNA l

XX
PS Example 5; SEQ ID NO 573; 99p

CC The invention relates to human
CC polynucleotides encoding them
XX

CC antibody or its antigen-binding
CC polypeptide of the invention,

CC
CC
CC
CC
CC

protein comprising a prostate invention are used to diagnose

CC the invention. Note: The sequence of the printed specification is

CC DIRECTLY FROM USPTO at SEGREGATED
XX
SO Sequence 132 AA.

Alignment Scores:

Score: 317.50
Percent Similarity: 61.1%

Best Local Similarity: 50.4%
Query Match: 34.2%

US-10-607-806-1 COPY 1 500 (1-500)

QY 431 GGCCTGAAGGATGCTAAGTGC
||| ...|||...

Db 4 GlyGluGlyGluAlaArgHis

1
 2
 3
 4
 5
 6
 7
 8
 9
 10
 11
 12
 13
 14
 15
 16
 17
 18
 19
 20
 21
 22
 23
 24
 25
 26
 27
 28
 29
 30
 31
 32
 33
 34
 35
 36
 37
 38
 39
 40
 41
 42
 43
 44
 45
 46
 47
 48
 49
 50
 51
 52
 53
 54
 55
 56
 57
 58
 59
 60
 61
 62
 63
 64
 65
 66
 67
 68
 69
 70
 71
 72
 73
 74
 75
 76
 77
 78
 79
 80
 81
 82
 83
 84
 85
 86
 87
 88
 89
 90
 91
 92
 93
 94
 95
 96
 97
 98
 99
 100
 101
 102
 103
 104
 105
 106
 107
 108
 109
 110
 111
 112
 113
 114
 115
 116
 117
 118
 119
 120
 121
 122
 123
 124
 125
 126
 127
 128
 129
 130
 131
 132
 133
 134
 135
 136
 137
 138
 139
 140
 141
 142
 143
 144
 145
 146
 147
 148
 149
 150
 151
 152
 153
 154
 155
 156
 157
 158
 159
 160
 161
 162
 163
 164
 165
 166
 167
 168
 169
 170
 171
 172
 173
 174
 175
 176
 177
 178
 179
 180
 181
 182
 183
 184
 185
 186
 187
 188
 189
 190
 191
 192
 193
 194
 195
 196
 197
 198
 199
 200
 201
 202
 203
 204
 205
 206
 207
 208
 209
 210
 211
 212
 213
 214
 215
 216
 217
 218
 219
 220
 221
 222
 223
 224
 225
 226
 227
 228
 229
 230
 231
 232
 233
 234
 235
 236
 237
 238
 239
 240
 241
 242
 243
 244
 245
 246
 247
 248
 249
 250
 251
 252
 253
 254
 255
 256
 257
 258
 259
 260
 261
 262
 263
 264
 265
 266
 267
 268
 269
 270
 271
 272
 273
 274
 275
 276
 277
 278
 279
 280
 281
 282
 283
 284
 285
 286
 287
 288
 289
 290
 291
 292
 293
 294
 295
 296
 297
 298
 299
 300
 301
 302
 303
 304
 305
 306
 307
 308
 309
 310
 311
 312
 313
 314
 315
 316
 317
 318
 319
 320
 321
 322
 323
 324
 325
 326
 327
 328
 329
 330
 331
 332
 333
 334
 335
 336
 337
 338
 339
 340
 341
 342
 343
 344
 345
 346
 347
 348
 349
 350
 351
 352
 353
 354
 355
 356
 357
 358
 359
 360
 361
 362
 363
 364
 365
 366
 367
 368
 369
 370
 371
 372
 373
 374
 375
 376
 377
 378
 379
 380
 381
 382
 383
 384
 385
 386
 387
 388
 389
 390
 391
 392
 393
 394
 395
 396
 397
 398
 399
 400
 401
 402
 403
 404
 405
 406
 407
 408
 409
 410
 411
 412
 413
 414
 415
 416
 417
 418
 419
 420
 421
 422
 423
 424
 425
 426
 427
 428
 429
 430
 431
 432
 433
 434
 435
 436
 437
 438
 439
 440
 441
 442
 443
 444
 445
 446
 447
 448
 449
 450
 451
 452
 453
 454
 455
 456
 457
 458
 459
 460
 461
 462
 463
 464
 465
 466
 467
 468
 469
 470
 471
 472
 473
 474
 475
 476
 477
 478
 479
 480
 481
 482
 483
 484
 485
 486
 487
 488
 489
 490
 491
 492
 493
 494
 495
 496
 497
 498
 499
 500
 501
 502
 503
 504
 505
 506
 507
 508
 509
 510
 511
 512
 513
 514
 515
 516
 517
 518
 519
 520
 521
 522
 523
 524
 525

Db 20 ValArgGlyGluThrAlaThrAsnPhePheLeuArgGlnGluSerGlyProValAla 39

QY 338 CAGGTGAGTGCAGTGTGCAATTTGGCTCAGTCACTGCAACCTCCACCTCCAGGTTCAAG 279

Db 40 GlnAlaGlyValGlnTrpHisAspLeuSerSerLeuGlnProLeuProHisArgPheLys 59

QY 278 TGATTCTTCTGCCCCAGCTCTCTGTAGCTGGGACTATAGGCAAGTCCACACACCTG 219

Db 60 GlnPheSerCysLeuSerLeuProHisSerTrpAspHisArgTyrAlaProProHisLeu 79

QY 218 GCTAATTTTGTATTTTATTATAGAGATAGTGTGTTTGGCCATGTTGGCCAGGCTGCTCAA 159

Db 80 AlalaPheCysSerPheSerArgaspGlyValSerLeuCysCysSerGlyTrpSerLys 99

QY 158 ACTCTGCTCTCAAGTGAATCCACCAACTCGGTCTCCAAAGTGTGTTGGGAATACAGACAT 99

Db 100 ThrProGlyLeuGlnGlnSerAlaCysLeuGlyLeuProLysCysTrpGlyTyrArgHis 119

QY 98 GAGCACCGCAGCTGGCTAATGAGTTACTATTA 66

Db 120 LysProProHisProAlaCysHisIleLeuLeu 130

RESULT 13

AAU01284

ID AAM01284 standard; protein; 135 AA.

AC AAM01284;

DT 04-OCT-2001 (first entry)

XX P789P splice variant contig 6 amino acid #2.

DE Human; prostate cancer; prostate-specific; diagnosis; vaccine;

KW cytostatic; gene therapy; metastasis.

XX Homo sapiens.

XX WO200151633-A2.

XX 19-JUL-2001.

XX 16-JAN-2001; 2001WO-US0001574.

XX 14-JAN-2000; 2000US-00483672.

XX (CORI-) CORIXA CORP.

XX Xu J, Dillon DC, Mitcham JL, Harlocker SL, Jiang Y, Reed SG;

PI Kalos MD, Fanger GR, Day CH, Retter MW, Stolk JA, Skeiky YAW;

PI Wang A, Meagher WJ;

XX WPI; 2001-425873/45.

XX New polynucleotide encoding a prostate-specific protein, for diagnosing,

PT monitoring and treating prostate cancer in a patient and for use in

PT vaccines.

XX Claim 2; Page 514-515; 543pp; English.

PS The present invention describes polynucleotide sequences (I) which encode

CC prostate-specific proteins (II). (I) and (II) have cytostatic activity,

CC and can be used in vaccine production and gene therapy. (I), (II),

CC antibodies to (II), fusion proteins comprising (II), and isolated T cells

CC prepared using (I) or (II) are used treat cancer in a patient. (I) and

CC the antibodies are also used in the detection of cancer in a patient. The

CC cancer that is diagnosed or treated is particularly prostate cancer. (I)

CC and (II) can be used in vaccines. The antibodies or (I) can be used for

CC monitoring the progression of cancer in a patient. (I) and (II) can also

CC be used to improve diagnostic and therapeutic methods for prostate

CC cancer. They can indicate the level of metastasis as well as the prostate

CC volume. AAH93357 to AAH93944 and AAM01115 to AAM01318 represent

CC polynucleotide and amino acid sequences used in the exemplification of

CC the present invention

XX SQ Sequence 135 AA;

Alignment Scores: 4.52e-29 Length: 135

Pred. No.: 317.50 Matches: 66

Score: 61.1% Conservative: 14

Percent Similarity: 50.4% Mismatches: 38

Best Local Similarity: 34.2% Indels: 13

Query Match: 2 Gaps: 2

DB: 4

US-10-607-806-1_COPY_1_500 (1-500) x AAM01284 (1-135)

QY 431 GGCCTGAAGATGCTAAGTGGTGCATTGGGCTGGCTCAAAAGAGAGAGACAGCTATGG 372

Db 4 GlyGluGlyGlnAlaArgHisValLeuHisGlyGlyArgGluArg----- 19

QY 371 TTGCAAGTAAAGTTATTAAATTACCTTTTTC----- 339

Db 20 ValArgGlyGluThrAlaThrAsnPhePheLeuArgGlnGluSerGlyProValAla 39

QY 338 CAGGTGAGTGCAGTGTGCAATTTGGCTCAGTCACTGCAACCTCCACCTCCAGGTTCAAG 279

Db 40 GlnAlaGlyValGlnTrpHisAspLeuSerSerLeuGlnProLeuProHisArgPheLys 59

QY 278 TGATTCTTCTGCCCCAGCTCTCTGTAGCTGGGACTATAGGCAAGTCCACACACCTG 219

Db 60 GlnPheSerCysLeuSerLeuProHisSerTrpAspHisArgTyrAlaProProHisLeu 79

QY 218 GCTAATTTTGTATTTTATTATAGAGATAGTGTGTTTGGCCATGTTGGCCAGGCTGCTCAA 159

Db 80 AlalaPheCysSerPheSerArgaspGlyValSerLeuCysCysSerGlyTrpSerLys 99

QY 158 ACTCTGCTCTCAAGTGAATCCACCAACTCGGTCTCCAAAGTGTGTTGGGAATACAGACAT 99

Db 100 ThrProGlyLeuGlnGlnSerAlaCysLeuGlyLeuProLysCysTrpGlyTyrArgHis 119

QY 98 GAGCACCGCAGCTGGCTAATGAGTTACTATTA 66

Db 120 LysProProHisProAlaCysHisIleLeuLeu 130

RESULT 14

AAU69929

ID AAU69929 standard; protein; 135 AA.

XX AAU69929;

AC AAU69929;

DT 30-JAN-2002 (first entry)

XX Human prostate cDNA encoded protein #74.

DE Human; prostate cancer; cytostatic; immunostimulant; tumour; immunogen.

XX Homo sapiens.

XX WO200173032-A2.

XX 04-OCT-2001.

XX 27-MAR-2001; 2001WO-US0009919.

XX 27-MAR-2000; 2000US-00536857.

PR 09-MAY-2000; 2000US-00586100.

PR 12-MAY-2000; 2000US-00570737.

PR 13-JUN-2000; 2000US-00593793.

PR 27-JUN-2000; 2000US-00605783.

PR 09-AUG-2000; 2000US-00636215.

PR 29-AUG-2000; 2000US-00651236.

PR 06-SEP-2000; 2000US-00657279.

PR 02-OCT-2000; 2000US-00679426.

PR 10-OCT-2000; 2000US-00685166.

PR 09-NOV-2000; 2000US-00709729.

XX

PA (CORI-) CORIXA CORP.
 XX Xu J, Dillon DC, Mitcham JL, Harlocker SL, Jiang Y, Kalos MD;
 PI Fanger GR, Retter MW, Stolk JA, Day CH, Vedvick TS, Carter D;
 PI Li SX, Wang A, Skeiky YAW, Hepler WT, Henderson RA;
 XX WPI; 2001-639232/73.
 DR N-PSDB; AAS64161.
 XX New human prostate-specific polypeptides and polynucleotides useful for
 PT the diagnosis and treatment of cancer, especially prostate cancer.
 PT Claim 2; Page 552; 579pp; English.
 XX The invention relates to isolated prostate-specific polynucleotides,
 CC polypeptides, fusion proteins of the polypeptides, antibodies raised
 CC against the polypeptides (or antigenic epitopes derived from them) and
 CC antigen-presenting cells expressing the polypeptides. The antibodies are
 CC useful for detecting the presence of cancer, especially prostate cancer.
 CC The polypeptides, polynucleotides and the antigen-presenting cells are
 CC useful for stimulating and/or expanding T cells specific for a tumour
 CC protein, and for inhibiting the development of cancer especially prostate
 CC cancer. Compositions comprising the polynucleotide and/or polypeptide are
 CC useful for stimulating an immune response, and for treating cancer. The
 CC oligonucleotide is useful for detecting cancer. The present sequence is a
 CC prostate specific polypeptide of the invention
 XX SQ Sequence 135 AA;
 Alignment Scores:
 Pred. No.: 4.52e-29 Length: 135
 Score: 317.50 Matches: 66
 Percent Similarity: 61.1% Conservative: 14
 Best Local Similarity: 50.4% Mismatches: 38
 Query Match: 34.2% Indels: 13
 DB: 4 Gaps: 2
 US-10-607-806-1_COPY_1_500 (1-500) x AAU69929 (1-135)
 Qy 431 GGCCTGAAGAGTGTAACTGGGTGCTTGGGGTGGCTCAAGAGGAGGACGATG 372
 Db 4 GlyGluGlyGluAlaArgHisValGluHisGlyGlyArgGluArg----- 19
 Qy 371 TTGCAAGTAAAGTTATTAAATACCTTTTTC----- 339
 Db 20 ValArgGlyGluThrAlaThrAsnPhePheLeuArgGlnGluSerGlyProValAla 39
 Qy 338 CAGCTGGAGTGCAGTGTGCAATCTGGCTACCTGCACTGCACCTCCACCTCCAGTTCAAG 279
 Db 40 GlnAlaGlyValGlnTrpHisAspLeuSerSerLeuGlnProLeuProHisArgPheLys 59
 Qy 278 TGATTCTTCTGCCCCAGCTCTGTGTAGCTGGCACTATAGCAAGTGCACACACCTG 219
 Db 60 GlnPheSerCysLeuSerLeuProHisSerTrpAspHisArgTyrAlaProProHisLeu 79
 Qy 218 GCTAATTTTCTATTATTTATTAGATAGTGTGTTTTCATGCTGGCCAGGCTGCTCAA 159
 Db 80 AlaAsnPheCysSerPheSerArgAspGlyValSerLeuCysSerGlyTrpSerLys 99
 Qy 158 ACTCCTGCTCTCAAGTATGATCCACCACTCGGTCTCCCAAGTCTCGGGAATACAGACAT 99
 Db 100 ThrProGlyLeuGlnGlnSerAlaCysLeuGlyLeuProLysCysTrpGlyTyrArgHis 119
 Qy 98 GAGCCACCGACCTGGCTAATGACTTACTATTA 66
 Db 120 LysProProHisProAlaCysHisGluLeuLeu 130
 RESULT 15
 ABU71820
 ID ABU71820 standard; protein; 135 AA.
 XX
 XX ABU71820;
 XX

DT 10-JUN-2003 (first entry)
 XX Prostate cancer specific antigen p776P #16.
 KW Prostate cancer; vaccine; gene therapy; cytostatic; fusion protein;
 KW immunogen; cancer; prostate specific antigen.
 XX Homo sapiens.
 OS Synthetic.
 XX US2002192763-A1.
 XX 19-DEC-2002.
 XX 29-JUN-2001; 2001US-00895793.
 PR 04-OCT-1999; 99US-0157455P.
 PR 04-OCT-2000; 2000US-00679272.
 PR 28-MAR-2001; 2001US-00822827.
 XX (XUJJ/) XU J.
 PA (DILL/) DILLON D C.
 PA (MITC/) MITCHAM J L.
 PA (HARL/) HARLOCKER S L.
 PA (JIAN/) JIANG Y.
 PA (KALO/) KALOS M D.
 PA (FANG/) FANGER G R.
 PA (RETT/) RETTER M W.
 PA (STOL/) STOLK J A.
 PA (DAYC/) DAY C H.
 PA (VEDV/) VEDVICK T S.
 PA (CART/) CARTER D.
 PA (LISX/) LI S X.
 PA (WANG/) WANG A.
 PA (SKEI/) SKEIKY Y A W.
 PA (HEPL/) HEPLER W T.
 PA (HEND/) HENDERSON R A.
 PA (HURA/) HURAL J.
 PA (MCNE/) MCNEILL P D.
 PA (HOUG/) HOUGHTON R L.
 PA (DBAS/) Y DE BASSOLS C V.
 PA (FOYT/) FOY T M.
 XX Xu J, Dillon DC, Mitcham JL, Harlocker SL, Jiang Y, Kalos MD;
 PI Fanger GR, Retter MW, Stolk JA, Day CH, Vedvick TS, Carter D;
 PI Li SX, Wang A, Skeiky YAW, Hepler WT, Henderson RA, Hural J;
 PI McNeill PD, Houghton RL, Y De BassolsCV, Foy TM;
 XX WPI; 2001-245062/25.
 DR N-PSDB; ACA59969.
 XX Prostate specific protein and its encoding polynucleotide, useful for the
 PT treatment and diagnosis of prostate cancer.
 PT Example 5; SEQ ID NO 884; 85pp; English.
 PS The invention describes a fusion protein comprising at least one amino
 XX acid sequence of immunogenic portions of any of the 3 sequences not
 CC defined in the specification, or sequences having at least 70 or 90 %
 CC sequence identity to any one of the 35 sequences defined in the USPTO web
 CC site, which is encoded by any of the 4 nucleotide sequences not defined
 CC in the specification. The fusion protein, composition and methods are
 CC useful for diagnosing, preventing and/or treating cancer, particularly
 CC prostate cancer. The proteins are useful as markers to indicate the
 CC presence or absence of cancer. This is the amino acid sequence of a
 CC prostate cancer specific antigen. Note: The sequence data for this patent
 CC did not form part of the printed specification, but was obtained in
 CC electronic format directly from the US patent office at
 CC seqdata.uspto.gov/sequence.html?DocID=US20020192763
 XX Sequence 135 AA;
 SQ Alignment Scores:

Search completed: February 8, 2006, 09:59:42
Job time : 86.529 secs

Db311PhePhePhePheGluThrGluSerHisSerValThrGlnAlaGlyValGlnTrpArg330

Qy317AATCTTGGCTCAGTCAACCTCCACCTCCAGGTTCAAGTGATTCTTCTGCCCCAGCCTC258

Db331AspLeuGlySerLeuGlnAlaProProProGlyPheMetProPheSerCysLeuSerLeu350

Qy257CTGTGTAGCTGGGACTATAGGCAAGTCCACACCTGGCGTAAT-----TTTGTATT204

Db351LeuArgThrTrpAspTyrArgArg-----ProHisHisAlaGlnLeuLeuPheCysIle368

Qy203TTTATTAGATAGTGTTCCTATGTCAGGCTGGCTCAAACTCTGCTCTCAAG144

Db369PheSerArgAsnGlyValLeuProCysTrpProGlyTrpSerArgThrProGlyLeu--M388

Qy143TGATCCACCAACTCGGCTCTCCAAAGTGTGGGAATACAGATGAGCCACCGACCT85

Db388etIleHisProLeuProPheProPheValLeuGlyLeuGlnAla**AlaThrAlaPro407

RESULT 2

F40201

artifact-warning sequence (translated ALU class F) - human

C:Species: Homo sapiens (man)

C:Date: 31-Mar-1992 #sequence_revision 11-Aug-1995 #text_change 19-May-2000

C:Accession: F40201

R:Claverie, J.M.

personal communication, 1992

A:Reference number: A40201

A:Accession: F40201

A:Molecule type: DNA

A:Residues: 1-673 <CLA>

A:Cross-references: UNIPARC:UPI000017CECD

R:Claverie, J.M.

Genomics 12, 838-841, 1992

A:Title: Identifying coding exons by similarity search: Alu-derived and other potentially

A:Reference number: A40200; PMID:92241891; PMID:1572661

A:Contents: annotation

C:Comment: This "warning" entry is a conceptual translation in all 6 reading frames of c in-frame stop codons are shown as 'X'.

C:Comment: Any significant similarity of a predicted protein sequence to a portion of th

Alignment Scores:

Pred. No.:4.86e-21Length:673

Score:268.00Matches:57

Percent Similarity:69.8%Conservative:3

Best Local Similarity:66.3%Mismatches:24

Query Match:28.9%Indels:2

DB:4Gaps:0

US-10-607-806-1_COPY_1_500 (1-500) x F40201 (1-673)

Qy339CCAGCTGGAGTGCAGTGGTGCATCTTGGCTCACTGCACCACTCCACCTCCAGGTTCAA280

Db589ProGlyTrpSerAlaVal-AlaSerPheGlySerLeuGlnProGlnProGlyPheLe508

Qy279GTGATTCTTCTGCCAGCTCTGTAGTGGGACTATAGGCAAGTGGCCACCACT220

Db608sArgPheSerCysLeuProCysSerTrpAspTyrArgHisGlyProProPheLe628

Qy219GGCTAAATTTTGTATTTTATTAGAGATAGTGTTCCTGATGGTGGCAGGCTGCTCA160

Db628uAlaAsnLeuCysIlePheAsnArgAspThrValSerProCys**SerGlyTrpSerG1648

Qy159AACTCTCTCTCAAGTATCCACCACTCGGCTCTCCCAAGTCTGGGAATACAGACA100

Db648nThrProAspLeuLys**SerAlaLeuLeuValSerGlnSerAlaGlyMetThrGlyV668

Qy99TGAGCCACCGCAC86

Db668aISerHisHisThr672

RESULT 4

C40201

artifact-warning sequence (translated ALU class C) - human

C:Species: Homo sapiens (man)

C:Date: 31-Mar-1992 #sequence_revision 11-Aug-1995 #text_change 19-May-2000

C:Accession: C40201

R:Claverie, J.M.

personal communication, 1992

A:Reference number: A40201

A:Accession: C40201

A:Molecule type: DNA

A:Residues: 1-613 <CLA>

A:Cross-references: UNIPARC:UPI000017CECA

R:Claverie, J.M.

Genomics 12, 838-841, 1992

A:Title: Identifying coding exons by similarity search: Alu-derived and other potential

A:Reference number: A40200; PMID:92241891; PMID:1572661

A:Contents: annotation

C:Comment: This "warning" entry is a conceptual translation in all 6 reading frames of c in-frame stop codons are shown as 'X'.

C:Comment: Any significant similarity of a predicted protein sequence to a portion of th

Alignment Scores:

Pred. No.:2.27e-21Length:673

Score:271.00Matches:58

Percent Similarity:74.4%Conservative:6

Best Local Similarity:67.4%Mismatches:20

Query Match:29.2%Indels:2

DB:4Gaps:0

US-10-607-806-1_COPY_1_500 (1-500) x F40201 (1-673)

Qy87GTGCGTGGCTCATGCTCTATTCAGCAGCTTGGGACCGAGTTGGTGATCACTT146

Db3ValTrpTrpLeuThrProValIleProAlaLeuTrpGluThrArgArg-AlaAspHisue22

Qy147GAGAGCAGGAGTTTGAGACGAGCTGGCCAAATGATGCAAAACACTATCTTAATAAAAT206

Db22uArgSerGlyVal**AspGlnProAspGlnHisGlyGluThrValSerLeuLeuLys142

Qy207ACAAAATAGCCAGTGTGGTGGCTTGTCTATATAGTCCAGCTACACAGGAGGTGG266

Db42eHisLysLeuAlaArgAsnGlyGlyProCysLeu**SerGlnLeuHisGlyArgGlnAr62

Qy267GCAGAGATCATCTGAACCTGGAGGTGCAGTGCAGTGCAGTGCAGTGCAGTGCAGT326

Db62gGlnGluAsnArgLeuAsnProGlyGly**GlyCysSerGluProLysLeu-AlaThrA82

Qy327CACTCCAGCTCGA340

Db82laLeuGlnProGly86

RESULT 3

Alignment Scores:
Pred. No.: 2,52e-20 Length: 613
Score: 261.50 Matches: 66
Percent Similarity: 52.5% Conservative: 8
Best Local Similarity: 46.8% Mismatches: 26
Query Match: 28.2% Indels: 41
DB: 4 Gaps: 5

US-10-607-806-1_COPY_1_500 (1-500) x C40201 (1-613)

```
QY 86 GTGGCGGCTCATCTGTTATCCAGCAGCTGGGAGACCGAGTGGGTGGATCACT 145
DB 208 GlyAlaValAlaHisAlaCysAsnProSerThrLeuGlyGlyGlyTyrPile--M 227
QY 146 TGAGAGCAGAGTTTGAGACCAAGCTGGCCACATGGCAAAACACTATCTCTAAATAAAA 205
DB 227 eArgProGlyValArgAspGlnProGlyGlnHisGlyLysThrProPheLeuLeuLysI 247
QY 206 TACAAAATA-----GCCAGGTGTGGTGGCACTTGCTATAGTCCAGCTTACACAGA 259
DB 247 leGlnLysIleSerTrpAlaTrpCysGlyArg-----Leu***SerHisValArgA 265
QY 260 GCTGGGGCAGAAATCACTTGAACCTGGGAGGTGGAGTGGCAGTGAAGCAAGATTGC 319
DB 265 rgLeuArgGlnGluAsnGlyIleAsnProGlyGlyAlaCysSerGluProArgSerA 285
QY 320 ACCACTGCACTCCAGCCTGG-----AAAAAAGG 349
DB 285 rgHisCysThrProAlaTrpValThrGlu***AspSerValSerLysLysLysL 305
QY 350 TAAATTAATAACTTACTTGCACCATAGCTGCTCTCTCTCTTTCAG----- 396
DB 305 ys*****Phe-----PhePhePhePhePheGluThrGluSerHis 321
QY 397 -----CCACCCCA 405
DB 321 erValThrGlnAlaGlyValGlnTrpArgAspLeuGlySerLeuGlnAlaProPro 340
RESULT 5
138022
hypothetical protein - human
C:Species: Homo sapiens (man)
C>Date: 01-Nov-1996 #sequence_revision 01-Nov-1996 #text_change 05-Nov-1999
C:Accession: I38022
R:Yang, S.S.; Zhang, K.; Vieira, W.; Taub, J.V.; Zeilstra-Ryalls, J.H.; Somerville, R.L.
Cancer Res. 50, 5658-5667, 1990
A:Title: A human hepatocellular carcinoma 3.0-kilobase DNA sequence transforms both rat
A:Reference number: I38021
A:Accession: I38022
A>Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: DNA
A:Residues: 1-196 <RES>
A:Cross-references: UNIPARC:UPI000011EIFB; EMBL:X55777; NID:g288143; PIDN:CAA39297.1; PI
```

Alignment Scores:
Pred. No.: 6,39e-18 Length: 196
Score: 239.50 Matches: 50
Percent Similarity: 63.2% Conservative: 10
Best Local Similarity: 52.6% Mismatches: 20
Query Match: 25.8% Indels: 15
DB: 2 Gaps: 1

US-10-607-806-1_COPY_1_500 (1-500) x I38022 (1-196)

```
QY 142 CACTTGAGCAGCAGGTTTGAGACCGCTGGCCCAACATGGCAAAACATATCTCTAATA 201
DB 3 HisLeuArgSerGlyValGlnAspTyrProGlyGlnHisGlyLysIleProSerLeuLeu 22
QY 202 AAAATACAAAATTAGCAGGTGGTGGCACTTGCCTTATAGTCCAGCTACACAGGAGG 261
DB 23 LysIleGlnGlnAlaGlyHisGlyGlyArgCysLeuGlnSerGlnLeuLeuArgArg 42
```

```
QY 262 CTGGGGCAGAGAAATCACTTGAACTGGAGGTGAGGTTGTCAGTGAGCCAGATTGCAC 321
DB 43 LeuArgGlnGlnAsnHisLeuAsnSerGlyArgGlyCysSerGluProLysSerHis 62
QY 322 CACTGCACTCCAGCCTGG----- 339
DB 63 LeuCysIleProAlaTrpValThrGluGlyAspSerValSerLysGlnAsnLysThrLys 82
QY 340 ---AAAAAAGGTAATAATAACTTTACTTGCACACCATAGCTGC 381
DB 83 AsnGluGlnHisLeuArgAsnAsnThrLysLysSerAsnSerCys 97
```

RESULT 6
A40201

C:Species: Homo sapiens (man)
C>Date: 31-Mar-1992 #sequence_revision 11-Aug-1995 #text_change 19-May-2000
C:Accession: A40201
R:Claverie, J.M.
personal communication, 1992
A:Reference number: A40201
A:Accession: A40201
A:Molecule type: DNA
A:Residues: 1-627 <CLA>
A:Cross-references: UNIPARC:UPI000017CEC8
R:Claverie, J.M.
Genomics 12, 838-841, 1992

A:Title: Identifying coding exons by similarity search: Alu-derived and other potential
A:Reference number: A40200; MUID:92241891; PMID:1572661
A:Contents: annotation
C:Comment: This "warning" entry is a conceptual translation in all 6 reading frames of
in-frame stop codons are shown as 'X'.
C:Comment: Any significant similarity of a predicted protein sequence to a portion of

Alignment Scores:
Pred. No.: 6,44e-16 Length: 627
Score: 221.50 Matches: 63
Percent Similarity: 52.5% Conservative: 11
Best Local Similarity: 44.7% Mismatches: 30
Query Match: 23.9% Indels: 37
DB: 4 Gaps: 7

US-10-607-806-1_COPY_1_500 (1-500) x A40201 (1-627)

```
QY 470 CATACAGACAGAGCAGGACCGACTGCTCTAGTATTAGGCTGAAGGATGCTAAGTGG 411
DB 291 HisSerSerLeuGlyAspArgAla--ArgLeuArgLeuGly----- 303
QY 410 GTGATTGGGGTGGCTCAAGAAGAGAGACGACGCTATGTTGCAAGTAAAG----- 360
DB 304 -----LysLys***LysGln-----LysIleLysLys***** 314
QY 359 -----TTATTAAATACCTTTTTTTC-----CAG 336
DB 315 ***PhePheThrPheLeuPheLeuPhePheSerGluThrGluSerArgSerValAlaLys 334
QY 335 GCTGGAGTGCAGTGGTGAATCTTGGCTCACTGCAACCTCCACCTCCAGGTTCAAGTGA 276
DB 335 AlaGlyValGlnTrpArgAspLeuGlySerLeuGlnAlaProProGlyPheThrPro 354
QY 275 TTCTTTCGCCCGCTCTGCTGTAGCTGGGACTATAGGCAAGTGCACACACCTGGCT 216
DB 355 PheSerCysLeuSerLeuProSerSerTrpAspTyrArgArg-----ProHisTyrAla 372
QY 215 AAT-----TTTTGTATTTTATTAGATAGTGTTTTGGCATGTTGGCCAGGCTGTCT 162
DB 373 ArgLeuIlePheCysIlePheSerArgAspGlyValSerProTrp***ProGlyTrpSer 392
QY 161 CAACACTCTGCTCTCAAGTGATCACCCCACTCGGTCTCCCAAGTGTCTGGGAATACAG 103
DB 393 ArgSerProAspLeu--ValIleArgProProTrpProProLysValLeuGlyLeuGln 411
```

RESULT 7
E40201
artifact-warning sequence (translated ALU class E) - human
C:Species: Homo sapiens (man)
C:Date: 31-Mar-1992 #sequence_revision 11-Aug-1995 #text_change 19-May-2000
C:Accession: E40201
R:Claverie, J.M.
personal communication, 1992
A:Reference number: A40201
A:Accession: E40201
A:Molecule type: DNA
A:Residues: 1-597 <CLA>
A:Cross-references: UNIPARC:UPI000017CEBC
R:Claverie, J.M.
Genomics 12, 838-841, 1992
A:Title: Identifying coding exons by similarity search: Alu-derived and other potential
A:Reference number: A40200; MUID:92241891; PMID:1572661
A:Contents: annotation
C:Comment: This "warning" entry is a conceptual translation in all 6 reading frames of c
in-frame stop codons are shown as 'X'.
C:Comment: Any significant similarity of a predicted protein sequence to a portion of th

Alignment Scores:
Pred. No.: 1,748-14 Length: 597
Score: 208.50 Matches: 52
Percent Similarity: 62.1% Conservative: 7
Best Local Similarity: 54.7% Mismatches: 24
Query Match: 22.5% Indels: 13
DB: 4 Gaps: 2

US-10-607-806-1_COPY_1_500 (1-500) x E40201 (1-597)

QY 347 CTTTTCCTCCAGCTGGAGTGGCGTGGTCAATCTTGGCTCACTGCAACCTCCACCTCCC 288
DB 510 LeuValAlaGlnAlaGlyValArgTrpHisAsnLeuThrAla-----523
QY 287 AGGTTCAAGTGAATTC-----TGCCCGAGCTCTCTGTGTAGTGGGAC 243
DB 524 AsnPheAlaSerTrpValGlnAlaLeuSerCysLeuSerLeuProSerSerTrpAsp 543
QY 242 TATAGCAAGTGGCCACACCTGGCTAATTTTGTATTTTATAGATAGTGTGTG 183
DB 544 TyrArgHisAlaProProArgProAlaAsn-PheIlePheLeuValGluMetGlyPheLe 563
QY 182 CCATGTGTGGCGGCTGCTCAAACTCTCTCAAGTGATCCACCCAA-CTCGGTCT 124
DB 563 whisValGlyGlnAlaGlyLeuLysLeuProThrSerGlyAspProProArgLeuGluLe 583
QY 123 CCCAAAGTCTGGGAATACAGACATGACGACCGCACCTGGCT 81
DB 583 uProLysAArgTrpAspTyrArgHisGluLeuProHisLeuAla 597

RESULT 8
A40201
artifact-warning sequence (translated ALU class A) - human
C:Species: Homo sapiens (man)
C:Date: 31-Mar-1992 #sequence_revision 11-Aug-1995 #text_change 19-May-2000
C:Accession: A40201
R:Claverie, J.M.
personal communication, 1992
A:Reference number: A40201
A:Accession: A40201
A:Molecule type: DNA
A:Residues: 1-627 <CLA>
A:Cross-references: UNIPARC:UPI000017CEC8
R:Claverie, J.M.
Genomics 12, 838-841, 1992
A:Title: Identifying coding exons by similarity search: Alu-derived and other potential
A:Reference number: A40200; MUID:92241891; PMID:1572661
A:Contents: annotation
C:Comment: This "warning" entry is a conceptual translation in all 6 reading frames of o

in-frame stop codons are shown as 'X'.
C:Comment: Any significant similarity of a predicted protein sequence to a portion of th

Alignment Scores:
Pred. No.: 2,55e-14 Length: 627
Score: 207.00 Matches: 48
Percent Similarity: 67.9% Conservative: 9
Best Local Similarity: 57.1% Mismatches: 25
Query Match: 22.3% Indels: 2
DB: 4 Gaps: 0

US-10-607-806-1_COPY_1_500 (1-500) x A40201 (1-627)

QY 89 GCGTGGCTCATGCTCTATTCCAGCACTTTGGAGACCGAGTTGGTGGTCACTTGA 148
DB 3 AlaArgSerHisAlaCysAsnProSerThrLeuGlyGlnGlyGlyArgIleThr--A 22
QY 149 GAGCAGGAGTTTGAGACGAGCCTGGCCAAACATGCAAAACACATCTCTTAATAAAATAC 208
DB 22 rGSerGlyAspArgAspHisProGlyTyrHisGlyGluThrProSerLeuLeuLysIleG 42
QY 209 AAAAATAGCAGGTGTGGGCACTTGCTCTATAGTCCAGCTACACAGGAGGCTGGGC 268
DB 42 InLysIleSerArgAla***CysGlyArgLeu***SerGlnLeuLeuGlyArgLeuArgG 62
QY 269 AGAAGATCATCTGAACCTGGAGGTGGAGTTGCAGTGAGCCAAAGATTGCACCACTGCA 328
DB 62 InGluAsnGlyValAsnProGlyGlyAlaCysSerGluProArgSerArgHisCysT 82
QY 329 CTCCAGCC 336
DB 82 hrProAla 84

RESULT 9

S65657
alpha-1C-adrenergic receptor splice form 2 - human
N:Alternate names: alpha-1C-adrenoceptor isoform 2
C:Species: Homo sapiens (man)
C:Date: 22-Nov-1996 #sequence_revision 22-Nov-1996 #text_change 09-Jul-2004
C:Accession: S65657; S65655
R:Tanaka, T.
submitted to the EMBL Data Library, July 1994
A:Reference number: S65656
A:Accession: S65657
A:Molecule type: mRNA
A:Residues: 1-499 <TAN>
A:Cross-references: UNIPROT:Q13675; UNIPARC:UPI000004E75E; EMBL:D32202; NID:g927208; PID:
R:Hirasawa, A.; Shibata, K.; Horie, K.; Takei, Y.; Obika, K.; Tanaka, T.; Muramoto, N.;
FEBS Lett. 363, 256-260, 1995
A:Title: Cloning, functional expression and tissue distribution of human alpha (1C)-adre
A:Reference number: S65654; MUID:95255557; PMID:7737411
A:Accession: S65655
A:Molecule type: mRNA
A:Residues: 424-499 <HIR>
A:Cross-references: UNIPARC:UPI00001778AB; EMBL:D32202
C:Genetics:
A:Gene: GDB:ADRA1C; ADRA1L1
A:Cross-references: GDB:128088; OMIM:104221
A:Map position: 8p21-8p11.2
C:Superfamily: vertebrate rhodopsin
C:Keywords: alternative splicing; neurotransmitter receptor

Alignment Scores:
Pred. No.: 2,76e-12 Length: 499
Score: 188.50 Matches: 43
Percent Similarity: 65.7% Conservative: 3
Best Local Similarity: 61.4% Mismatches: 19
Query Match: 20.3% Indels: 5
DB: 2 Gaps: 1

US-10-607-806-1_COPY_1_500 (1-500) x S65657 (1-499)

artifact-warning sequence (translated ALU class B) - human

C;Species: Homo sapiens (man)

C;Date: 31-Mar-1992 #sequence_revision 11-Aug-1995 #text change 19-May-2000

C;Accession: B40201

R;Claverie, J.M.

personal communication, 1992

A;Reference number: A40201

A;Accession: B40201

A;Molecule type: DNA

A;Residues: 1-301 <CLA>

A;Cross-references: UNIPARC:UPI000017CEC9

R;Claverie, J.M.

Genomics 12, 838-841, 1992

A;Title: Identifying coding exons by similarity search: Alu-derived and other potential

A;Reference number: A40200; MUID:92241891; PMID:1572661

A;Contents: annotation

C;Comment: This "warning" entry is a conceptual translation in all 6 reading frames of c
c;concentr: annotation
in-frame stop codons are shown as 'X'.

C;Comment: Any significant similarity of a predicted protein sequence to a portion of the

[illegible]

Alignment Scores:

Pred. No.:	6.19e-09	Length:	301
Score:	158.00	Matches:	42
Percent Similarity:	58.8%	Conservative:	15
Best Local Similarity:	43.3%	Mismatches:	31
Query Match:	17.0%	Indels:	9
DB:	4	Gaps:	2

```

Pled. NO.: 6.19E-09
Score: 158.00
Length:
Matches:

```

Percent Similarity:

Percent Similarity: 38.8%
Best Local Similarity: 43.3%
Mismatches: 38.8%
Conservative: 43.3%

Best Local Similarity:	43.3%
Query Match:	17.0%
Mismatches:	9
Indels:	9

Query MATCH:
DB:

UB: 4
Gap: 2

US-10-607-806-1 COPY 1 500 (1-500) x B40201 (1-301)

QY 362 AAGTTATTAAATACCTTTTTTCCAGGCTGGAGTGCAGTGGTGC AATCTTGGCTCACTG 303

[illegible]

Db 205 GlnLeuIleLeuPheTyrPheTyrArgAlaGlyValLeuLeuCysProSerTrpSer 224

Qy 302 CAACCTCCACCTCCACGGTTCAAGTGATTCTTCTGCCCCAGCCCTC

Db 225 **ThrProGlyLeuLysGlnSerSerTyrPhe-----SerIle

1000

QY 242 TATAGGCAAGTGCACACCTGGCTAAT-----TTTGTATTTT 201

Db 243 TyrArgCys**ProProCysProAla*****ProSer*****

1. The first part of the document is a list of names, each followed by a number in parentheses. The names are: "John (1)", "Mary (2)", "Peter (3)", "Paul (4)", "David (5)", "Michael (6)", "James (7)", "Robert (8)", "Richard (9)", "Thomas (10)", "Christopher (11)", "Daniel (12)", "Matthew (13)", "Anthony (14)", "Mark (15)", "Steven (16)", "Andrew (17)", "Kenneth (18)", "Edward (19)", "George (20)", "William (21)", "Charles (22)", "Joseph (23)", "Thomas (24)", "Christopher (25)", "Daniel (26)", "Matthew (27)", "Anthony (28)", "Mark (29)", "Steven (30)", "Andrew (31)", "Kenneth (32)", "Edward (33)", "George (34)", "William (35)", "Charles (36)", "Joseph (37)", "Thomas (38)", "Christopher (39)", "Daniel (40)", "Matthew (41)", "Anthony (42)", "Mark (43)", "Steven (44)", "Andrew (45)", "Kenneth (46)", "Edward (47)", "George (48)", "William (49)", "Charles (50)", "Joseph (51)", "Thomas (52)", "Christopher (53)", "Daniel (54)", "Matthew (55)", "Anthony (56)", "Mark (57)", "Steven (58)", "Andrew (59)", "Kenneth (60)", "Edward (61)", "George (62)", "William (63)", "Charles (64)", "Joseph (65)", "Thomas (66)", "Christopher (67)", "Daniel (68)", "Matthew (69)", "Anthony (70)", "Mark (71)", "Steven (72)", "Andrew (73)", "Kenneth (74)", "Edward (75)", "George (76)", "William (77)", "Charles (78)", "Joseph (79)", "Thomas (80)", "Christopher (81)", "Daniel (82)", "Matthew (83)", "Anthony (84)", "Mark (85)", "Steven (86)", "Andrew (87)", "Kenneth (88)", "Edward (89)", "George (90)", "William (91)", "Charles (92)", "Joseph (93)", "Thomas (94)", "Christopher (95)", "Daniel (96)", "Matthew (97)", "Anthony (98)", "Mark (99)", "Steven (100)".

200 ATTAGAGATAGTGTTTTGCCCATGTTGGCCAGGCTGGTCTCAAACCT

252

Db
263 Ile-GluLeuGlySerCysTyrValAlaGlnAlaGlyLeuGluLeuValSerSerAs 282

Search completed: February 8, 2006, 10:03:29

Job time : 23.4044 secs

THIS PAGE BLANK (USPTO)

GenCore version 5.1.7
Copyright (c) 1993 - 2006 Bioacceleration Ltd.

OM nucleic - protein search, using frame_plus_n2p model

Run on: February 8, 2006, 09:18:47 ; Search time 24.2672 Seconds
(without alignments)
2907.338 Million cell updates/sec

Title: US-10-607-806-1_COPY_1_500
Perfect score: 928
Sequence: 1 gaccactactegactttgtg.....accacataagaggactga 500

Scoring table:
BLOSUM62
Xgapop 10.0 , Xgapext 0.5
Ygapop 10.0 , Ygapext 0.5
Fgapop 6.0 , Fgapext 7.0
Delop 6.0 , Delext 7.0

Searched: 2166443 seqs, 705528306 residues

Total number of hits satisfying chosen parameters: 4322886

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Command line parameters:

-MODEL=frame+ n2p.model -DEV=xlp
-O=/abs/ABSSWEB_epool/US10607806/runat_08022006_091819_26843/app_query.fasta.1
-DB=uniprot -QFMT=fastan -SUFFIX=n2p.rup -MINMATCH=0.1 -LOOPEXT=0
-UNITS=bits -START=1 -END=-1 -MATRIX=blosum62 -TRANS=human40.cdi -LIST=45
-DOCALIGN=200 -THR SCORE=pct -THR MAX=100 -THR MIN=0 -ALIGN=15 -MODE=LOCAL
-OUTFMT=ptc -NORM=ext -HEAPSIZE=500 -MINLEN=0 -MAXLEN=200000000 -HOST=abs02p
-USR=US10607806 @CGN 1.1 694 @runat_08022006_091819_26843 -NCPU=6 -ICPU=3
-NO MMAP -NEG SCORES=0 -WAIT -DSPLOCK=100 -LONGLOG -DEV TIMEOUT=120
-WARN TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5 -FGAPOP=6 -FGAPEXT=7
-YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database : Uniprot_05.80.*
1: uniprot_sprot.*
2: uniprot_trembl.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
C 1	359	38.7	593	1 ALU7_HUMAN	P39194 homo sapien
C 2	343	37.0	593	1 ALU6_HUMAN	P39193 homo sapien
C 3	322	34.3	593	1 ALU7_HUMAN	P39194 homo sapien
C 4	318	34.3	260	2 Q6ZQR8_HUMAN	O6zqr8 homo sapien
C 5	317	34.2	136	2 Q96NR6_HUMAN	O96nr6 homo sapien
C 6	314	33.8	158	2 Q8NAL9_HUMAN	O8nal9 homo sapien
C 7	311.5	33.6	138	2 Q8N874_HUMAN	O8n874 homo sapien
C 8	311	33.5	140	2 Q6ZP99_HUMAN	O6zpp9 homo sapien
C 9	310	33.4	124	2 Q6ZP96_HUMAN	O6zvf6 homo sapien
C 10	307.5	33.1	202	2 Q6ZUA4_HUMAN	O6zuu4 homo sapien
C 11	307	33.1	585	1 ALU5_HUMAN	P39192 homo sapien
C 12	305.5	32.9	122	2 Q6ZVH4_HUMAN	O6zv44 homo sapien
C 13	304	32.8	593	1 ALU6_HUMAN	P39193 homo sapien
C 14	301.5	32.5	166	2 Q6ZUC0_HUMAN	O6zuc0 homo sapien
C 15	301	32.4	591	1 ALU1_HUMAN	P39188 homo sapien
C 16	300	32.3	94	2 Q8NED3_HUMAN	O8ned3 homo sapien

C 17	297	32.0	171	2	Q9H728_HUMAN	Q9h728 homo sapien
C 18	297	32.0	591	1	ALU1_HUMAN	P39188 homo sapien
C 19	293.5	31.6	591	1	ALU8_HUMAN	P39195 homo sapien
C 20	293	31.6	160	2	Q6ZWD2_HUMAN	O6zwd2 homo sapien
C 21	293	31.6	218	2	Q8NAC7_HUMAN	O8nac7 homo sapien
C 22	290	31.2	165	2	Q6ZS53_HUMAN	O6zss3 homo sapien
C 23	290	31.2	165	2	Q6ZT71_HUMAN	O6zt71 homo sapien
C 24	289	31.1	115	2	Q9N083_MACFA	Q9n083 macaca fasc
C 25	288.5	31.1	137	2	Q6ZS97_HUMAN	O6zsr7 homo sapien
C 26	286	30.8	127	2	Q6ZVW2_HUMAN	O6zvw2 homo sapien
C 27	285	30.7	130	2	Q6ZSV2_HUMAN	O6zsv2 homo sapien
C 28	284.5	30.7	147	2	Q6ZUA2_HUMAN	O6zuu2 homo sapien
C 29	284.5	30.7	591	1	ALU8_HUMAN	P39195 homo sapien
C 30	284	30.6	220	2	Q5SVL1_HUMAN	O5svl1 homo sapien
C 31	284	30.6	587	1	ALU2_HUMAN	P39189 homo sapien
C 32	283	30.5	138	2	Q6ZUF3_HUMAN	O6zuf3 homo sapien
C 33	282	30.4	124	2	Q6ZU63_HUMAN	O6zu63 homo sapien
C 34	282	30.4	585	1	ALU5_HUMAN	P39192 homo sapien
C 35	281	30.3	127	2	Q9H743_HUMAN	O9h743 homo sapien
C 36	281	30.3	138	2	Q6ZU05_HUMAN	O6zuu5 homo sapien
C 37	281	30.3	150	2	Q6ZPA0_HUMAN	O6zpa0 homo sapien
C 38	279.5	30.1	125	2	Q6ZTEL_HUMAN	O6ztel homo sapien
C 39	279	30.1	156	2	Q8NBH4_HUMAN	O8nbh4 homo sapien
C 40	279	30.1	164	2	Q6ZUK0_HUMAN	O6zuk0 homo sapien
C 41	278.5	30.0	137	2	Q6ZP21_HUMAN	O6zpp21 homo sapien
C 42	278	30.0	118	2	Q9H387_HUMAN	O9h387 homo sapien
C 43	278	30.0	143	2	Q6ZVQ4_HUMAN	O6zvc4 homo sapien
C 44	277	29.8	124	2	Q6ZSN0_HUMAN	O6zsn0 homo sapien
C 45	277	29.8	132	2	Q6ZNG3_HUMAN	O6zng3 homo sapien

ALIGNMENTS

RESULT 1
ALU7_HUMAN
ID ALU7_HUMAN STANDARD; PRT; 593 AA.
AC P39194;
DT 01-FEB-1995 (Rel. 31, Created)
DT 01-FEB-1995 (Rel. 31, Last sequence update)
DT 10-MAY-2005 (Rel. 47, Last annotation update)
DE Alu subfamily SQ sequence contamination warning entry.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini; Hominidae;
OC Homo.
OX NCBI_TaxID=9606;
RN [1]
RP NUCLEOTIDE SEQUENCE.
RX MEDLINE=95021758; PubMed=7935834; DOI=10.1038/371752a0;
RA Claverie J.-M., Makalowski W.;
RT "Alu alert";
RL Nature 371:752-752(1994).
RN [2]
RP CONCEPT.
RX MEDLINE=92241891; PubMed=1572661;
RA Claverie J.-M.;
RT "Identifying coding exons by similarity search: alu-derived and other
potentially misleading protein sequences.";
RL Genomics 12:838-841(1992).
RN [3]
RP ALU FAMILIES CLASSIFICATION.
RX MEDLINE=88333009; PubMed=3138422;
RA Quentin Y.;
RT "The Alu family developed through successive waves of fixation closely
connected with primate lineage history.";
RL J. Mol. Evol. 27:194-202(1988).
RN [4]
RP ALU FAMILIES CLASSIFICATION
RX MEDLINE=91178815; PubMed=1706781;
RA Jurka J., Milosavljevic A.;
RT "Reconstruction and analysis of human Alu genes.";
RL J. Mol. Evol. 32:105-121(1991).
CC -!- MISCELLANEOUS: Various analyses indicate that Alu repeats fall

CC into 8 subfamilies (Ref.3 and Ref.4). Therefore, 8 Alu warning
CC consensus sequences have been constituted that contain all six
CC frames conceptual translations of each of these classes of Alu
CC repeats.
CC -!- MISCELLANEOUS: Isolated 'X' indicates the presence of a stop
CC codon, 'XXX' is used to separate the various translation phases.
CC -!- CAUTION: This Alu entry is provided in order to avoid the further
CC pollution of protein sequence databases with Alu-derived amino
CC acid sequences.
CC -!- CAUTION: Alu repetitive sequences are interspersed in human and
CC primate genomes with an average spacing of 4 kb. Some of them are
CC actively transcribed by pol III. Normal transcripts may contain
CC Alu-derived sequences in 5' or 3' untranslated regions. However,
CC cDNA libraries also contain partial and/or rearranged cDNAs
CC ligated with Alu-derived sequence in any orientation. Although Alu
CC elements (especially situated on the complementary strand) have a
CC great potential to create additional/alternative exons,
CC consideration should be given to the possibility that the presence
CC of an Alu in an open reading frame may have resulted from a
CC cloning artifact or may be due to misinterpretation of sequencing
CC data. This point has been overlooked on several occasions, with
CC the consequence of erroneous Alu-derived amino acid sequences
CC being reported.
CC -!- CAUTION: Any significant similarity of a putative protein sequence
CC with an Alu-translated entry must be taken as a warning that a
CC part of Alu repeat may have been artifactually included in the
CC coding nucleotide sequence.
CC -----
CC This Swiss-Prot entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use as long as its content is in no way modified and this statement is not
CC removed.
CC -----
CC EMBL: U14573; -!- NOT ANNOTATED_CDS; Genomic_DNA.
CC InterPro: IPR002218; GIDA.
CC ProDom: PD003738; GIDA; 1.
CC KW Hypothetical protein.
CC FT REGION 1 97 Frame-1.
CC FT REGION 101 196 Frame-2.
CC FT REGION 200 295 Frame-3.
CC FT REGION 299 395 Frame-4.
CC FT REGION 399 494 Frame-5.
CC FT REGION 498 593 Frame-6.
CC SQ SEQUENCE 593 AA; 64455 MW; 54A4F50F33A6089F CRC64;

CC Alignment Scores:
CC Pred. No.: 5,18e-30 Length: 593
CC Score: 359.00 Matches: 72
CC Percent Similarity: 66.1% Conservative: 12
CC Best Local Similarity: 56.7% Mismatches: 37
CC Query Match: 38.7% Indels: 6
CC DB: Gaps: 2

CC US-10-607-806-1_COPY_1_500 (1-500) x ALU7_HUMAN (1-593)
CC QY 461 ACAGAGACAGGCACTGCTCTAGTCTTTAGGCGCTGAGGAGTCTAAGTGGGTGTTGGG 402
CC Db 275 SerArgAspArgAlaThrAlaLeuGlnProGlyGlnGlnGluArgAsnSerVal----- 292
CC QY 401 GGTGGCTCAAGAAGGAGAGACGACTATGCTGTCAGAGTAAAGTATTATTACCTTTT 342
CC Db 293 -----SerTysLys*****PhePheLeu-----ArgArgSerPheAlaLeuVal 308
CC QY 341 TTCAGGCTGGAGTGGTGCATCTTGGCTCACTGCACCTCCACTCCAGGTTTC 282
CC Db 309 AlaGlnAlaGlyValGlnTrpArgAspLeuGlySerLeuGlnProProProGlyPhe 328
CC QY 281 AAGTGATTTCTTGCAGGCTCCCTGCTGTAGCTGGGACTATAGGCAAGTGCACACAC 222
CC Db 329 LysArgPheSerCysLeuSerLeuProSerSerTipAspTyrArgArgProProArg 348
CC QY 221 CTGGCTAAATTTTGTATTTTATTAGAGATAGTGTGTTTGGCCAGGCTGGTCT 162

Db 349 ProAlaAsnPheCysIlePheSerArgAspGlyValSerProCysTrpProGlyTyrSer 368
CC QY 161 CAAACTCTCTGCTCTCAAGTCATCCACCCCACTCGGTCTCCCAAGTGGTGGGAATACAGA 102
CC Db 369 ArgThrProAspLeuArg***SerThrArgLeuGlyLeuProLysCysTrpAspTyrArg 388
CC QY 101 CATGAGCCACCGCACCTGGCT 81
CC Db 389 ArgGluProProArgProAla 395

CC RESULT 2
CC ID ALU6_HUMAN STANDARD; PRT; 593 AA.
CC AC P39193; 1995 (Rel. 31, Created)
CC DT 01-FEB-1995 (Rel. 31, Last sequence update)
CC DT 01-FEB-1995 (Rel. 31, Last annotation update)
CC DT 10-MAY-2005 (Rel. 47, Last annotation update)
CC DE Alu subfamily SP sequence contamination warning entry.
CC OS Homo sapiens (Human)
CC OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
CC OC Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini; Hominidae;
CC OC Homo.
CC OX NCBI_TaxID=9606;
CC RN [1]
CC RP NUCLEOTIDE SEQUENCE.
CC RA MEDLINE=95021758; PubMed=7935834; DOI=10.1038/371752a0;
CC RX Claverie J.-M., Makalowski W.;
CC RT "Alu alert."
CC RL Nature 371:752-752 (1994).
CC RN [2]
CC RP CONCEPT.
CC RX MEDLINE=92241891; PubMed=1572661;
CC RA Claverie J.-M.;
CC RT "Identifying coding exons by similarity search: alu-derived and other
CC RT potentially misleading protein sequences."
CC RL Genomics 12:838-841 (1992).
CC RN [3]
CC RP ALU FAMILIES CLASSIFICATION.
CC RX MEDLINE=89333009; PubMed=3138422;
CC RA Quentin Y.;
CC RT "The Alu family developed through successive waves of fixation closely
CC RT connected with primate lineage history."
CC RL J. Mol. Evol. 27:194-202 (1988).
CC RN [4]
CC RP ALU FAMILIES CLASSIFICATION.
CC RX MEDLINE=91178815; PubMed=1706781;
CC RA Jurka J., Milosavljevic A.;
CC RT "Reconstruction and analysis of human Alu genes."
CC RL J. Mol. Evol. 32:105-121 (1991).
CC CC -!- MISCELLANEOUS: Various analyses indicate that Alu repeats fall
CC into 8 subfamilies (Ref.3 and Ref.4). Therefore, 8 Alu warning
CC consensus sequences have been constituted that contain all six
CC frames conceptual translations of each of these classes of Alu
CC repeats.
CC -!- MISCELLANEOUS: Isolated 'X' indicates the presence of a stop
CC codon, 'XXX' is used to separate the various translation phases.
CC -!- CAUTION: This Alu entry is provided in order to avoid the further
CC pollution of protein sequence databases with Alu-derived amino
CC acid sequences.
CC -!- CAUTION: Alu repetitive sequences are interspersed in human and
CC primate genomes with an average spacing of 4 kb. Some of them are
CC actively transcribed by pol III. Normal transcripts may contain
CC Alu-derived sequences in 5' or 3' untranslated regions. However,
CC cDNA libraries also contain partial and/or rearranged cDNAs
CC ligated with Alu-derived sequence in any orientation. Although Alu
CC elements (especially situated on the complementary strand) have a
CC great potential to create additional/alternative exons,
CC consideration should be given to the possibility that the presence
CC of an Alu in an open reading frame may have resulted from a
CC cloning artifact or may be due to misinterpretation of sequencing
CC data. This point has been overlooked on several occasions, with
CC the consequence of erroneous Alu-derived amino acid sequences

CC being reported.
CC -! CAUTION: Any significant similarity of a putative protein sequence
CC with an Alu-translated entry must be taken as a warning that a
CC part of Alu repeat may have been artifactually included in the
CC coding nucleotide sequence.
CC -----
CC This Swiss-Prot entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use as long as its content is in no way modified and this statement is not
CC removed.
CC -----
DR EMBL; U14572; -; NOT_ANNOTATED_CDS; Genomic_DNA.
DR InterPro; IPR002218; GIDA.
DR ProDom; PD003738; GIDA; 1.
KW Hypothetical protein.
FT REGION 1 97
FT REGION 101 196
FT REGION 200 295
FT REGION 299 395
FT REGION 399 494
FT REGION 498 593
SQ SEQUENCE 593 AA; 64640 MW; 136EF344AACD12A2 CRC64;

Alignment Scores:
Pred. No.: 2.94e-28 Length: 593
Score: 343.00 Matches: 69
Percent Similarity: 63.8% Conservative: 12
Best Local Similarity: 54.3% Mismatches: 40
Query Match: 37.0% Indels: 6
DB: 1 Gaps: 2

US-10-607-806-1_COPY_1_500 (1-500) x ALU6_HUMAN (1-593)
QY 461 ACAGAGGACGAGGACTCTCTAGTTTAGCCTGAGGATGCTAAGTGGTGTGTTGGG 402
Db 275 SerArgAspArgAlaIleAlaLeuGlnProGlyGlnGlnGluArgSerVal----- 292
QY 401 GGTGGCTCAAGAAGGAGGAGCAGCTATGTTTGCAGTAAAGTATTATTAATACCTTTT 342
Db 293 -----SerLysLys*****PheLeu-----ArgArgSerPheAlaLeuVal 308
QY 341 TTCAGGCTGAGTGCAGTGTGCAATCTTGGCTACCTGCACCTCCACCTCCAGGTTTC 282
Db 309 AlaGlnAlaGlyValGlnTrpArgAspLeuGlySerProGlnProProProGlyPhe 328
QY 281 AAGTGATCTTCTCCCGAGCTCTCTGTAGCTGGGACTATAGCAGTCCACACAC 222
Db 329 LysArgPheSerCysLeuSerLeuProSerSerTrpAspTyrArgHisAlaProProArg 348
QY 221 CTGGCTAATTTTGTATTATTTATAGATAGTGTGTTTGCATGTTGGCCAGGCTGTCT 162
Db 349 ProAlaSerPheCysIlePheSerArgAspGlyValSerProCysTrpSerGlyTrpSer 368
QY 161 CAAGCTCTCTCAAGTATGATCCACCACTCGGTCTCCCAAGTCTGGGAATACAGA 102
Db 369 ArgThrProAspLeuArg***SerAlaArgLeuGlyLeuProLysCysTrpAspTyrArg 388
QY 101 CATGAGCCACGCGCTGGCT 81
Db 389 ArgGluProProArgProAla 395

RESULT 3
ID ALU7_HUMAN STANDARD; PRT; 593 AA.
AC P39194;
DT 01-FEB-1995 (Rel. 31, Last sequence update)
DT 01-FEB-1995 (Rel. 31, Last sequence update)
DT 10-MAY-2005 (Rel. 47, Last annotation update)
DE Alu subfamily SQ sequence contamination warning entry.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini; Hominiidae;

OC Homo.
OX NCBI_TaxID=9606;
RN [1]
RP NUCLEOTIDE SEQUENCE.
EX MEDLINE=95021758; PubMed=7935834; DOI=10.1038/371752a0;
RA Claverie J.-M.; Makalowski W.;
RT "Alu alert."
RL Nature 371:752-752 (1994).
RN [2]
RP CONCEPT.
EX MEDLINE=92241891; PubMed=1572661;
RA Claverie J.-M.;
RT "Identifying coding exons by similarity search: alu-derived and other
RT potentially misleading protein sequences."
RL Genomics 12:838-841 (1992).
RN [3]
RP ALU FAMILIES CLASSIFICATION.
EX MEDLINE=88333009; PubMed=3138422;
RA Quent in Y.;
RT "The Alu family developed through successive waves of fixation closely
RT connected with primate lineage history."
RL J. Mol. Evol. 27:194-202 (1988).
RN [4]
RP ALU FAMILIES CLASSIFICATION.
EX MEDLINE=91178815; PubMed=1706781;
RA Jurka J.; Milosavljevic A.;
RT "Reconstruction and analysis of human Alu genes."
RL J. Mol. Evol. 32:105-121 (1991)
CC -!- MISCELLANEOUS: Various analyses indicate that Alu repeats fall
CC into 8 subfamilies (Ref.3 and Ref.4). Therefore, 8 Alu warning
CC consensus sequences have been constituted that contain all six
CC frames conceptual translations of each of these classes of Alu
CC repeats.
CC -!- MISCELLANEOUS: Isolated 'X' indicates the presence of a stop
CC codon, 'XXX' is used to separate the various translation phases.
CC -!- CAUTION: This Alu entry is provided in order to avoid the further
CC pollution of protein sequence databases with Alu-derived amino
CC acid sequences.
CC -!- CAUTION: Alu repetitive sequences are interspersed in human and
CC primate genomes with an average spacing of 4 kb. Some of them are
CC actively transcribed by pol III. Normal transcripts may contain
CC Alu-derived sequences in 5' or 3' untranslated regions. However,
CC cDNA libraries also contain partial and/or rearranged cDNAs
CC ligated with Alu-derived sequence in any orientation. Although Alu
CC elements (especially situated on the complementary strand) have a
CC great potential to create additional/alternative exons,
CC consideration should be given to the possibility that the presence
CC of an Alu in an open reading frame may have resulted from a
CC cloning artifact or may be due to misinterpretation of sequencing
CC data. This point has been overlooked on several occasions, with
CC the consequence of erroneous Alu-derived amino acid sequences
CC being reported.
CC -!- CAUTION: Any significant similarity of a putative protein sequence
CC with an Alu-translated entry must be taken as a warning that a
CC part of Alu repeat may have been artifactually included in the
CC coding nucleotide sequence.
CC -----
CC This Swiss-Prot entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use as long as its content is in no way modified and this statement is not
CC removed.
CC -----
DR EMBL; U14573; -; NOT_ANNOTATED_CDS; Genomic_DNA.
DR InterPro; IPR002218; GIDA.
DR ProDom; PD003738; GIDA; 1.
KW Hypothetical protein.
FT REGION 1 97
FT REGION 101 196
FT REGION 200 295
FT REGION 299 395
FT REGION 399 494
FT REGION 498 593

SQ SEQUENCE 593 AA; 64455 MW; 54A4F50F33A6089F CRC64;

Alignment Scores:
Pred. No.: 5.9e-26 Length: 593
Score: 322.00 Matches: 60
Percent Similarity: 76.1% Conservative: 7
Best Local Similarity: 68.2% Mismatches: 21
Query Match: 34.7% Indels: 0
DB: Gaps: 0

US-10-607-806-1_COPY_1_500 (1-500) x ALU7_HUMAN (1-593)

QY 84 CAGTGGGTGCTCATGTCATTTCCAGCACCTTTGGGACCGAGTGTTGGTGATCA 143
::: |||||
Db 2 ArgAlaArgTrpLeuThrProValIleProAlaLeuTrpGluAlaGlySer 21
::: |||||
QY 144 CTGGAGCAGGAGGTTTGAGACCAGCGCTGGCCAAACATGCATAATCTTAATAA 203
|||||
Db 22 ProGluValArgSerArgProAlaTaTProThrTriP**AsnProValSerThrLys 41
|||||
QY 204 AATCAAAAATTAGCCAGGTGGTGGCGACTTGCCTATAGTCCCAGCTACACAGAGGCT 263
|||||
Db 42 AsnThrLysIleSerArgAlaTrpTrpArgAlaProValIleProAlaThrArgGluAla 61
|||||
QY 264 GGSGCAGAACAATCACATTGAACCTGGGAGGTGGAGGTTCGAGTGGAGCCAAAGATTGCCA 323
|||||
Db 62 GluAlaGlyGluSerLeuGluProGlyArgArgLeuGln**AlaGluIleAlaPro 81
|||||
QY 324 CTGCATCCAGCCTGGAAAAAAG 347
|||||
Db 82 LeuHisSerLeuGlyAsnLys 89
|||||

RESULT 4

Q6ZQR8 HUMAN PRELIMINARY; PRT; 260 AA.

ID Q6ZQR8 HUMAN PRELIMINARY;
AC Q6ZQR8;
DT 05-JUL-2004 (TrEMBLrel. 27, Created)
DT 05-JUL-2004 (TrEMBLrel. 27, Last sequence update)
DT 05-JUL-2004 (TrEMBLrel. 27, Last annotation update)
DE Hypothetical protein FLJ45973.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini; Hominoidea;
OC Homo.
OX NCBI_TaxID=9606;
RN [1]
RP NUCLEOTIDE SEQUENCE.
RC TISSUE=Placenta;
RA Takahashi H., Yamazaki K., Watanabe K., Kumagai A., Itakura S.,
RA Fukuzumi Y., Fujimori Y., Komiyama M., Sugiyama T., Irie R.,
RA Otsuki T., Sato H., Wakamatsu A., Ishii S., Yamamoto J., Isono Y.,
RA Kawai-Hio Y., Saito K., Nishikawa T., Kimura K., Yamashita H.,
RA Matsuo K., Nakamura Y., Sekine M., Kikuchi H., Kanda K., Wagatsuma M.,
RA Murakami K., Kanehori K., Tanaka H., Tanaka H., Tanaka H., Sugiyama A.,
RA Kawakami B., Suzuki Y., Sugano S., Nagahara K., Masuhara Y., Nagai K.,
RA Isoqai T.;
RL Submitted (JUL-2003) to the EMBL/GenBank/DDBJ databases.
DR EMBL: AK128808; BAC87615.1; -; mRNA.
SQ SEQUENCE 260 AA; 28819 MW; 5D5C34F441BC5994 CRC64;

Alignment Scores:
Pred. No.: 1.34e-25 Length: 260
Score: 318.00 Matches: 61
Percent Similarity: 63.1% Conservative: 4
Best Local Similarity: 59.2% Mismatches: 16
Query Match: 34.3% Indels: 2
DB: Gaps: 2

US-10-607-806-1_COPY_1_500 (1-500) x Q6ZQR8_HUMAN (1-260)

QY 338 CAGGCTGGAGTGCAGTGGTGGCAATCTGGCTCACTGCACCTCCACCTCCAGGTTCAAG 279
::: |||||
Db 2 ArgAlaArgTrpLeuThrProValIleProAlaLeuTrpGluAlaGlySer 21
::: |||||

RA Sugawara M., Takahashi M., Kanda K., Yokoi T., Furuya T., Kikkawa E.,
RA Omura Y., Abe K., Kaminara K., Katsuta N., Sato K., Tanikawa M.,
RA Yamazaki M., Ninomiya K., Ishibaashi T., Yamashita H., Murakawa K.,
RA Fujimori K., Tanai H., Kimata M., Watanabe M., Hiraoka S., Chiba Y.,
RA Ishida S., Ono Y., Takiguchi S., Watanabe S., Yosida M., Hotuta T.,
RA Kusano J., Kanehori K., Takahashi-Fujii A., Hara H., Tanase T.-O.,
RA Nomura Y., Togiya S., Komai F., Hara R., Takeuchi K., Arita M.,
RA Imose N., Musashino K., Yuuki H., Oshima A., Sasaki N., Aotsuka S.,
RA Yoshikawa Y., Matsunawa H., Ichihara T., Shiohata N., Sano S.,
RA Moriya S., Moniyama H., Satoh N., Takami S., Terashima Y., Wakebe H.,
RA Hishigaki H., Watanabe T., Sugiyama A., Takemoto M., Kawakami B.,
RA Yamazaki M., Watanabe K., Kumagai A., Itakura S., Fukuzumi Y.,
RA Fujimori K., Komiyama M., Tashiro H., Tanigami A., Fujiwara T.,
RA Ono T., Yamada K., Fujii Y., Ozaki K., Hirao M., Ohmori Y.,
RA Kawabata A., Hikiji T., Kobatake N., Inagaki H., Ikema Y., Okamoto S.,
RA Okitani R., Kawakami T., Nozuchi S., Itoh T., Shigeta K., Senba T.,
RA Matsumura K., Nakajima Y., Mizuno T., Morinaga M., Sasaki M.,
RA Togashi T., Oyama M., Hata H., Watanabe M., Komatsu T., Nakagawa K.,
RA Mizushima-Sugano J., Satoh T., Shirai Y., Takahashi Y., Yamashita R.,
RA Okumura K., Nagase T., Nomura N., Kikuchi H., Masuho Y., Yamashita R.,
RA Nakai K., Yada T., Nakamura Y., Ohara O., Isogai T., Sugano S.;
RT "Complete sequencing and characterization of 21,243 full-length human
RT cDNAs.";
RL Nat. Genet. 36:40-45(2004).
DR EMBL; AK097214; BAC04976.1; -; mRNA.
DR Ensembl; ENSG0000198424; Homo sapiens.
SQ SEQUENCE 138 AA; 15427 MW; 685645DF92E31994 CRC64;

Alignment Scores:
Pred. No.: 5,96e-25 Length: 138
Score: 311.50 Matches: 65
Percent Similarity: 70.7% Conservative: 5
Best Local Similarity: 65.7% Mismatches: 27
Query Match: 33.6% Indels: 2
DB: 2 Gaps: 1

US-10-607-806-1_COPY_1_500 (1-500) x Q8N874_HUMAN (1-138)
QY 377 CTATGGTTCAGTAAGATTATTAATTCACCTTTTCCAGCTGGAGTCAGTGGTGC 318
Db 12 LeuPheLeuPheLeuSerTrpSerPheThrLeuValAlaGlnThrGlyValGlnTrpArg 31
QY 317 AATCTTGGCTCACTGCAACTCCACTCCAGCTCCAGTGAATCTTCTGCCCGCCCTC 258
Db 32 AsnLeuGlySerLeuGlnProProProGlyPheLeuArgPheSerCysLeuGlyLeu 51
QY 257 CTGTGTAGCTGGACTATAGGCAAGTCCACACACCTGGCTAAATTTTGTATTTATT 198
Db 52 ProSerSerTrpAspTrpArg-His---HisHisAlaTrpLeuPheValLeuLeuAl 70
QY 197 AGAGATAGTGTCTTGGCATGTGGCCAGGCTGCTCAAACTCCTCTCAAGTGATCC 138
Db 70 aGluMetGlyPheCysHisValGlyGlnAlaGlyLeuGluLeuLeuThrSerSerGluPr 90
QY 137 ACCCAACTCGGTCTCCAAAGTCTGGGATACAGACATGAGCCACGCGACTGG 83
Db 90 oProAlaSerAlaSerGlnSerAlaGlyIleThrGlyMetSerHisHisThrTrp 108

RESULT 8
Q6ZP99_HUMAN
ID Q6ZP99_HUMAN PRELIMINARY; PRT; 140 AA.
AC Q6ZP99;
DT 05-JUL-2004 (TrEMBLrel. 27, Created)
DT 05-JUL-2004 (TrEMBLrel. 27, Last sequence update)
DT 05-JUL-2004 (TrEMBLrel. 27, Last annotation update)
DE Hypothetical protein FLJ26180.
OS Homo sapiens (Human)
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Euthera; Euarchontoglires; Primates; Catarrhini; Hominidae;
OC Homo.
OX NCBI_TaxID=9606;
RN [1]

RP NUCLEOTIDE SEQUENCE.
RC TISSUE=Adrenal gland;
RA Oshima A., Takahashi-Fujii A., Tanase T., Imose N., Takeuchi K.,
RA Arita M., Musashino K., Yuuki H., Hara H., Suzuki Y., Hata H.,
RA Nakagawa K., Mizuno S., Morinaga M., Kawamura M., Sugiyama T.,
RA Irie R., Otsuki T., Sato H., Nishikawa T., Sugiyama A., Kawakami B.,
RA Nagai K., Isogai T., Sugano S.;
RL Submitted (JUL-2003) to the EMBL/GenBank/DBJ databases.
DR EMBL; AK129691; BAC85220.1; -; mRNA.
SQ SEQUENCE 140 AA; 15350 MW; 27A3F9D2C23FF9DD CRC64;

Alignment Scores:
Pred. No.: 6,79e-25 Length: 140
Score: 311.00 Matches: 60
Percent Similarity: 72.2% Conservative: 5
Best Local Similarity: 66.7% Mismatches: 25
Query Match: 33.5% Indels: 0
DB: 2 Gaps: 0

US-10-607-806-1_COPY_1_500 (1-500) x Q6ZP99_HUMAN (1-140)
QY 352 TTACCTCTTTTCCAGCTGGAGTCGAGTCGATCTTGGCTACTGCACTCCACCTCCAC 293
Db 30 LeuAlaLeuLeuProLysLeuGluCysArgGlyThrIleSerAlaHisCysAsnLeuHis 49
QY 292 CTCCAGGTTCAAGTGATCTTCTGCCCCAGCTCTGTGTAGCTGGAGCTATAGGCAAG 233
Db 50 LeuProGlySerSerSerAppPheProAlaSerAlaSerGlnValAlaGlyThrGlyAla 69
QY 232 TGCCACCACACCTGGCTAAATTTTGTATTTTATTAGAGATAGTGTGTTGTCATGTTGGC 173
Db 70 CysHisHisAlaTrpLeuIlePheValPheLeuValGluAlaGlyPheHisHisValGly 89
QY 172 CAGGCTGTCTCAAACTCTCTCTCAAGTATCCACCACTCGGTCTCCCAAGTGCT 113
Db 90 GlnAspGlyLeuGluLeuLeuThrSerAsnAspProThrLeuAlaSerGlnSerAla 109
QY 112 GGGATACAGACATGAGCCACGCGACTGG 83
Db 110 GlyIleThrGlyValSerHisArgAlaTrp 119

RESULT 9
Q6ZVP6_HUMAN
ID Q6ZVP6_HUMAN PRELIMINARY; PRT; 124 AA.
AC Q6ZVP6;
DT 05-JUL-2004 (TrEMBLrel. 27, Created)
DT 05-JUL-2004 (TrEMBLrel. 27, Last sequence update)
DT 05-JUL-2004 (TrEMBLrel. 27, Last annotation update)
DE Hypothetical protein FLJ42633.
OS Homo sapiens (Human)
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Euthera; Euarchontoglires; Primates; Catarrhini; Hominidae;
OC Homo.
OX NCBI_TaxID=9606;
RN [1]

NUCLEOTIDE SEQUENCE.
RP TISSUE=cerebellum;
RA Ninomiya K., Wagatsuma M., Kanda K., Kondo H., Yokoi T., Kodaira H.,
RA Furuya T., Takahashi M., Kikkawa E., Omura Y., Abe K., Kaminara K.,
RA Katsuta N., Sato K., Tanikawa M., Yamazaki M., Sugiyama T., Irie R.,
RA Otsuki T., Sato H., Wakamatsu A., Ishii S., Yamamoto J., Isono Y.,
RA Kawai-Hio Y., Saito K., Nishikawa T., Kimura K., Yamashita H.,
RA Matsuo K., Nakamura Y., Sekine M., Kikuchi H., Murakawa K.,
RA Kanehori K., Takahashi-Fujii A., Oshima A., Sugiyama A., Kawakami B.,
RA Suzuki Y., Sugano S., Nagahari K., Masuho Y., Nagai K., Isogai T.;
RL Submitted (JUL-2003) to the EMBL/GenBank/DBJ databases.
DR EMBL; AK124624; BAC85905.1; -; mRNA.
DR GO; GO:0004872; F:receptor activity; IEA.
DR InterPro; IPR000152; ASX_hydroxyl_S.
DR PROSITE; PS00010; ASX_HYDROXYL; UNKNOWN_1.
DR Receptor.
SQ SEQUENCE 124 AA; 14173 MW; 108AA891087A31B4 CRC64;

Alignment Scores:

Pred. No.: 8,49e-25 Length: 124
 Score: 310.00 Matches: 59
 Percent Similarity: 80.0% Conservatives: 5
 Best Local Similarity: 73.8% Mismatches: 14
 Query Match: 33.4% Indels: 2
 DB: 2 Gaps: 0

US-10-607-806-1_COPY_1_500 (1-500) x Q6ZVF6_HUMAN (1-124)

QY 338 CAGGCTGGAGTGCAGTGGTCAATCTTGCTCACTGCAACCTCCACCTCCAGGTTCAAG 279
 Db 46 GlnAlaGlyValGlnTrpCysAspLeuSerLeuGlnProLeuProGlyPheLeu 65
 QY 278 TGATCTCTTCTGCCAGCTCTCTGTAGTGGACTATAGGCAAGTGCACCAACCTG 219
 Db 66 GlnPhePheCysLeuSerLeuSerLeuSerTrpAspTyrArgGlnLeuProCysLeu 85
 QY 218 GCTAATTTTGTATTTTATTAGATAGTGTTCCTGTCATGTTGGCCAGCTGGTCTCAA 159
 Db 86 AlaAsnPheCysIlePheSerArgAspSerValSerProCysTrpSerGlyTrpSerGln 105
 QY 158 ACTCTGCTCTCAAGTGTACCACTCCCACTCGTCTCCCAAGTGTGGGAATACAG 103
 Db 106 ThrProAspLeu--LeuLeuArgProProArgProProLeuValLeuGlyLeuGln 123

RESULT 10

Q6ZUA4_HUMAN
 ID Q6ZUA4_HUMAN PRELIMINARY; PRT; 202 AA.
 AC Q6ZUA4
 DT 05-JUL-2004 (TrEMBLrel. 27, Created)
 DT 05-JUL-2004 (TrEMBLrel. 27, Last sequence update)
 DT 05-JUL-2004 (TrEMBLrel. 27, Last annotation update)
 DE Hypothetical protein FLJ43866.
 OS Homo sapiens (Human)
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini; Homiidae;
 OC Homo.
 OC NCBI_TaxID=9606;
 RN [1]
 RP NUCLEOTIDE SEQUENCE.
 RC TISSUE-Testis;
 RA Tashiro H., Yamazaki M., Watanabe K., Kumagai A., Itakura S.,
 RA Fukuzumi Y., Fujimori Y., Komiyama M., Sugiyama T., Irie R.,
 RA Otsuki T., Sato H., Wakamatsu A., Ishii S., Yamamoto J., Isono Y.,
 RA Kawai-Hio Y., Saito K., Nishikawa T., Kimura K., Yamashita H.,
 RA Matsuo K., Nakamura Y., Sekine M., Kikuchi H., Kanda K., Wagatsuma M.,
 RA Murakawa K., Kanehori K., Takahashi-Fujii A., Oshima A., Sugiyama A.,
 RA Kawakami B., Suzuki Y., Sugano S., Nagahara K., Masuho Y., Nagai K.,
 RA Isozaki T.,
 RL Submitted (JUL-2003) to the EMBL/GenBank/DBJ databases.
 DR EMBL; AK125854; BAC86322.1; -; mRNA.
 SQ SEQUENCE 202 AA; 22509 MW; 7C402B7E86DD01FA CRC64;

Alignment Scores:

Pred. No.: 1.79e-24 Length: 202
 Score: 307.50 Matches: 68
 Percent Similarity: 66.4% Conservatives: 5
 Best Local Similarity: 61.8% Mismatches: 26
 Query Match: 33.1% Indels: 11
 DB: 2 Gaps: 2

US-10-607-806-1_COPY_1_500 (1-500) x Q6ZUA4_HUMAN (1-202)

QY 391 AGAGGAGAACAGTATGTTGCAAGTAAAGTATTATTAATTCCTTTTTC----- 338
 Db 30 ArgSerGlnAspSerTyrAsn-----ProPheIleLeuLeuPheSerArgArg 46
 QY 337 -----AGGCTGAGTGCAGTGTGCAATCTTGGCTCACTGCAACCTC 296
 Db 47 SerLeuSerLeuSerProArgLeuGluCysSerGlyValIleLeuAlaHisCysAsnLeu 66
 QY 295 CACCTCCAGGTTCAAGTGTATCTTCTGCCCCCAGCTCTGTGTAGCTGGGACTATAGGC 236

Db 67 HisLeuProGlySerSerAspSerProAlaSerAlaSerArgValAlaGlyIleThrVal 86
 QY 235 AAGTCCCAACCCACCTGGCTAATTTTCTATTTTATTAGATAGTGTGTTGCCATGTT 176
 Db 87 AlaArgHisHisAlaTrpLeuIlePheValPheLeuValGluThrGlyPheHisVal 106
 QY 175 GGCACGGCTGGTCTCAAACTCTCTCAAGTGTATCCACCACTCGGTCTCCCAAAGT 116
 Db 107 GlyGlnAlaGlyLeuGluLeuLeuThrSerGlyAspProThrSerAlaSerGlnSer 126
 QY 115 GCTGGGAATACAGATCATGAGCCACCGACC 86
 Db 127 AlaGlyIleThrGlyValSerHisThr 136

RESULT 11
 ALU5_HUMAN
 ID ALU5_HUMAN STANDARD; PRT; 585 AA.
 AC P39192;
 DT 01-FEB-1995 (Rel. 31, Created)
 DT 01-FEB-1995 (Rel. 31, Last sequence update)
 DT 10-MAY-2005 (Rel. 47, Last annotation update)
 DE Alu subfamily SC sequence contamination warning entry.
 OS Homo sapiens (Human)
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini; Homiidae;
 OC Homo.
 OC NCBI_TaxID=9606;
 RN [1]
 RP NUCLEOTIDE SEQUENCE.
 RA Claverie J.-M.; Makalowski W.;
 RT "Alu alert."
 RL Nature 371:752-752 (1994).
 RN [2]
 RP MEDLINE=92241891; PubMed=1572661;
 RA Claverie J.-M.;
 RT "Identifying coding exons by similarity search: alu-derived and other
 RT potentially misleading protein sequences.";
 RL Genomics 12:838-841 (1992).
 RN [3]
 RP ALU FAMILIES CLASSIFICATION.
 RA Quentín Y.;
 RT "The Alu family developed through successive waves of fixation closely
 RT connected with primate lineage history.";
 RL J. Mol. Evol. 27:194-202 (1988).
 RN [4]
 RP ALU FAMILIES CLASSIFICATION.
 RA MEDLINE=91178815; PubMed=1706781;
 RA Jurka J., Milosavljevic A.;
 RT "Reconstruction and analysis of human Alu genes.";
 RL J. Mol. Evol. 32:105-121 (1991).
 CC -/- MISCELLANEOUS: Various analyses indicate that Alu repeats fall
 CC into 8 subfamilies (Ref.3 and Ref.4). Therefore, 8 Alu warning
 CC consensus sequences have been constituted that contain all six
 CC frames conceptual translations of each of these classes of Alu
 CC repeats.
 CC -/- MISCELLANEOUS: Isolated 'X' indicates the presence of a stop
 CC codon, 'XXX' is used to separate the various translation phases.
 CC -/- CAUTION: This Alu entry is provided in order to avoid the further
 CC pollution of protein sequence databases with Alu-derived amino
 CC acid sequences.
 CC -/- CAUTION: Alu repetitive sequences are interspersed in human and
 CC primate genomes with an average spacing of 4 kb. Some of them are
 CC actively transcribed by pol III. Normal transcripts may contain
 CC Alu-derived sequences in 5' or 3' untranslated regions. However,
 CC cDNA libraries also contain partial and/or rearranged cDNAs
 CC ligated with Alu-derived sequence in any orientation. Although Alu
 CC elements (especially situated on the complementary strand) have a
 CC great potential to create additional/alternative exons,
 CC consideration should be given to the possibility that the presence

CC of an Alu in an open reading frame may have resulted from a
 CC cloning artifact or may be due to misinterpretation of sequencing
 CC data. This point has been overlooked on several occasions, with
 CC the consequence of erroneous Alu-derived amino acid sequences
 CC being reported.
 CC -!- CAUTION: Any significant similarity of a putative protein sequence
 CC with an Alu-translated entry must be taken as a warning that a
 CC part of Alu repeat may have been artifactually included in the
 CC coding nucleotide sequence.

CC This Swiss-Prot entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use as long as its content is in no way modified and this statement is not
 CC removed.

CC EMBL; U14571; ; NOT_ANNOTATED_CDS; Genomic_DNA.

KW Hypothetical protein.
 FT REGION 1 95 Frame-1.
 FT REGION 99 193 Frame-2.
 FT REGION 197 291 Frame-3.
 FT REGION 295 389 Frame-4.
 FT REGION 393 487 Frame-5.
 FT REGION 491 585 Frame-6.
 SQ SEQUENCE 585 AA; 63994 MW; 462E8C4F493650A7 CRC64;

Alignment Scores:
 Pred. No.: 2,6e-24 Length: 585
 Score: 307.00 Matches: 61
 Percent Similarity: 72.8% Conservativeness: 6
 Best Local Similarity: 66.3% Mismatches: 23
 Query Match: 33.1% Indels: 2
 DB: 1 Gaps: 0

US-10-607-806-1_COPY_1_500 (1-500) x ALU5_HUMAN (1-585)

QY 353 ATTACCTTTTTCAGAGTGGAGTGGTGAATCTTGGCTACTGCAACTCA 294
 Db ValSerLeuCysArgGlnAlaGlyValGlnTrpArgAspLeuGlySerLeuGlnProPro 417
 QY 293 CTTCCAGGTTCAGAGTTCCTTCTGCCAGCTCCTGTAGTGGAGTATAGGCAA 234
 Db ProProGlyPheLysArgPheSerCysLeuSerLeuProSerSerTrpAspTrpArgArg 437
 QY 233 GTGCCACACACCTGGCTAATTTTGTATTTTATTTAGAGATAGTGTGGTGG 174
 Db AlaProProArgProAlaAsnPheCysIlePheSerArgAspGlyValSerProCysTrp 457
 QY 173 CGAGGTGGTCTCAAACTCTGCTCAAGTGTATCCCACTCGGTCTCCCAAGTGC 114
 Db ProGlyTrpSerArg--SerLeuAspLeuValIleArgProProArgProProLysVal 477
 QY 113 TGGGATACAGACATGAGCCACCGACCTGGC 82
 Db euGlyLeuGlnAla***AlaThrAlaProGly 487

RESULT 12

Q6ZVH4_HUMAN
 ID Q6ZVH4_HUMAN PRELIMINARY; PRT; 122 AA.
 AC Q6ZVH4; HUMAN
 DT 05-JUL-2004 (TrEMBLrel. 27, Created)
 DT 05-JUL-2004 (TrEMBLrel. 27, Last sequence update)
 DE Hypothetical protein FLJ42572.
 OS Homo sapiens (human).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini; Hominidae;
 OC Homo.
 OX NCBI_TaxID=9606;
 RN [1]
 RP NUCLEOTIDE SEQUENCE.
 RC TISSUE=Cerebellum;
 RA Ninomiya K., Wagatsuma M., Kanda K., Kondo H., Yokoi T., Kodaira H.,

RA Furuya T., Takahashi M., Kikkawa E., Omura Y., Abe K., Kamihara K.,
 RA Katsuta N., Sato K., Tanikawa M., Yamazaki M., Sugiyama T., Irie R.,
 RA Katsuki T., Sato H., Wakamatsu A., Ishii S., Yamamoto J., Isono Y.,
 RA Kawai-Hio Y., Saito K., Nishikawa T., Kimura K., Yamashita H.,
 RA Matsuo K., Nakamura Y., Sekine M., Kikuchi H., Murakawa K.,
 RA Kanehori K., Takahashi-Fujii A., Oshima A., Sugiyama A., Kawakami B.,
 RA Suzuki Y., Sugano S., Nagahari K., Masuho Y., Nagai K., Isogai T.,
 RL Submitted (JUL-2003) to the EMBL/GenBank/DBJ databases.
 DR EMBL; AK124563; BAC85887.1; ; mRNA.
 SQ SEQUENCE 122 AA; 13547 MW; 39472EF24DED96D3 CRC64;

Alignment Scores:
 Pred. No.: 2,64e-24 Length: 122
 Score: 305.50 Matches: 63
 Percent Similarity: 65.3% Conservativeness: 1
 Best Local Similarity: 64.3% Mismatches: 24
 Query Match: 32.9% Indels: 10
 DB: 2 Gaps: 1

US-10-607-806-1_COPY_1_500 (1-500) x Q6ZVH4_HUMAN (1-122)

QY 347 CTTTTTTC-----CAGGTGAGTGCAGTGGTCAAT 315
 Db LeuPhePheIleGluThrGluSerArgSerValAlaGlnAlaGlyValGlnTrpCysAsp 44
 QY 314 CTTGGCTCAGTGCACCTCCACCTCCAGGTTCAGTGTCTTCTGCCAGCTCCTCG 255
 Db LeuSerSerLeuGlnProProGlyPheLysArgPheSerCysLeuSerLeuPro 64
 QY 254 TGTAGCTGGAGTATAGGCAAGTGCACACACCTGGCTGCTAATTTTGTATTTTATTA 195
 Db SerSerTrpAspTrpArgCysProProArgLeuValThrPheCysIlePheSerArg 84
 QY 194 GATAGTGTTCCTGAGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGG 135
 Db AspArgValSerSerCysTrpProGly-GlyLeuLysLeuLeuThrSerGlyAspLeuPr 104
 QY 134 CAATCGCTCTCCCAAGTGTGGGAATACAGACATGAGCCACCTCGG 83
 Db AlaSerAlaSerGlnSerAlaGlyValThrGlyMetSerHisThrTrp 121

RESULT 13

ALU6_HUMAN
 ID ALU6_HUMAN STANDARD; PRT; 593 AA.
 AC P39193;
 DT 01-FEB-1995 (Rel. 31, Created)
 DT 01-FEB-1995 (Rel. 31, Last sequence update)
 DT 10-MAY-2005 (Rel. 47, Last annotation update)
 DE Alu subfamily SP sequence contamination warning entry.
 OS Homo sapiens (human).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini; Hominidae;
 OC Homo.
 OX NCBI_TaxID=9606;
 RN [1]
 RP NUCLEOTIDE SEQUENCE.
 RX MEDLINE=95021758; PubMed=7935834; DOI=10.1038/371752a0;
 RA Claverie J.-M., Makalowski W.;
 RT "Alu alert."
 RL Nature 371:752-752 (1994).
 RN [2]
 RP CONCEPT.
 RX MEDLINE=92241891; PubMed=1572661;
 RA Claverie J.-M.;
 RT "Identifying coding exons by similarity search: alu-derived and other
 RT potentially misleading protein sequences.";
 RL Genomics 12:838-841 (1992).
 RN [3]
 RP ALU FAMILIES CLASSIFICATION.
 RX MEDLINE=88333009; PubMed=3138422;
 RA Quentin Y.;
 RT "The Alu family developed through successive waves of fixation closely
 RT connected with primate lineage history.";

RL J. Mol. Evol. 27:194-202(1988).
 RN [4]
 RP ALU FAMILIES CLASSIFICATION.
 RX MEDLINE=9117815; PubMed=1706781;
 RA Jurka J., Milosavljevic A.;
 RT "Reconstruction and analysis of human Alu genes.";
 RL J. Mol. Evol. 32:105-121(1991).
 CC -!- MISCELLANEOUS: Various analyses indicate that Alu repeats fall
 CC into 8 subfamilies (Ref.3 and Ref.4). Therefore, 8 Alu warning
 CC consensus sequences have been constituted that contain all six
 CC frames conceptual translations of each of these classes of Alu
 CC repeats.
 CC -!- MISCELLANEOUS: Isolated 'X' indicates the presence of a stop
 CC codon, 'XXX' is used to separate the various translation phases.
 CC -!- CAUTION: This Alu entry is provided in order to avoid the further
 CC pollution of protein sequence databases with Alu-derived amino
 CC acid sequences.
 CC -!- CAUTION: Alu repetitive sequences are interspersed in human and
 CC primate genomes with an average spacing of 4 kb. Some of them are
 CC actively transcribed by pol III. Normal transcripts may contain
 CC Alu-derived sequences in 5' or 3' untranslated regions. However,
 CC cDNA libraries also contain partial and/or rearranged cDNAs
 CC ligated with Alu-derived sequence in any orientation. Although Alu
 CC elements (especially situated on the complementary strand) have a
 CC great potential to create additional/alternative exons,
 CC consideration should be given to the possibility that the presence
 CC of an Alu in an open reading frame may have resulted from a
 CC cloning artifact or may be due to misinterpretation of sequencing
 CC data. This point has been overlooked on several occasions, with
 CC the consequence of erroneous Alu-derived amino acid sequences
 CC being reported.
 CC -!- CAUTION: Any significant similarity of a putative protein sequence
 CC with an Alu-translated entry must be taken as a warning that a
 CC part of Alu repeat may have been artifactually included in the
 CC coding nucleotide sequence.
 CC -----
 CC This Swiss-Prot entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use as long as its content is in no way modified and this statement is not
 CC removed.
 CC -----
 CC EMBL; U14572; -; NOT_ANNOTATED_CDS; Genomic_DNA.
 CC InterPro; IPR002218; GIDA.
 CC ProDom; PD003738; GIDA; 1.
 CC Hypothetical protein.
 CC REGION 1 97 Frame-1.
 CC REGION 101 196 Frame-2.
 CC REGION 200 295 Frame-3.
 CC REGION 299 395 Frame-4.
 CC REGION 399 494 Frame-5.
 CC REGION 498 593 Frame-6.
 CC SEQUENCE 593 AA; 64640 MW; 136EF344AACD12A2 CRC64;
 Alignment Scores:
 Pred. No.: 5,55e-24 Length: 593
 Score: 304.00 Matches: 60
 Percent Similarity: 74.4% Conservative: 4
 Best Local Similarity: 69.8% Mismatches: 22
 Query Match: 32.8% Indels: 0
 DB: 1 Gaps: 0
 US-10-607-806-1_COPY_1_500 (1-500) x ALU6_HUMAN (1-593)
 QY 91 GTGGGCTCATGTCGTGATTCAGACACTTGGGAGACCGAGTTGGGTGGATCACTTGA 150
 Db 104 GlyGlySerArgLeu***SerGlnHisPheGlyArgProArgGalaAspHisLeuArg 123
 QY 151 GCAGGAGCTTGGACACCGCTGGCCACATCGGCAAAACACTATCTCTAATAAATAACAA 210
 Db 124 SerGlyValArgAspGlnProAspGlnHisGlyGluThrProSerLeuLeuIleGln 143
 QY 211 AAATTAGCCAGGTGTGTGGTGCATTCCTGCTATAGTCCAGCTACACAGGAGGCTGGGCGAG 270

Db 144 LysLeuAlaGlyArgGlyGlyAlaCysLeu***SerGlnLeuLeuGlyArgLeuArgGln 163
 QY 271 AAGAAATCACTTGAACCTGGAGGTGGAGTTGCAGTGCAGCAAGATTCCACCACTGCCT 330
 Db 164 GlnAsnArgLeuAenProGlyGlyGlyGlyGlyGlyGlyGlyGlyGlyGlyGlyGly 183
 QY 331 CCAGCTGGGAAAAAAGG 348
 Db 184 ProAlaTrpAlaThrArg 189
 RESULT 14
 Q6ZUC0 HUMAN
 ID Q6ZUC0 HUMAN PRELIMINARY; PRT; 166 AA.
 AC Q6ZUC0;
 DT 05-JUL-2004 (TrEMBLrel. 27, Created)
 DT 05-JUL-2004 (TrEMBLrel. 27, Last sequence update)
 DE Hypothetical protein FLJ43828.
 OS Homo sapiens (Human).
 CC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
 CC Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini; Homnidae;
 CC Homo.
 CC NCBI_TaxID=9606;
 RN [1]
 RP NUCLEOTIDE SEQUENCE.
 RC TISSUE=Testis;
 RA Ota T., Nakagawa S., Senoh A., Mizuguchi H., Inagaki H., Sugiyama T.,
 RA Irie K., Otsuki T., Sato H., Wakamatsu A., Ishii S., Yamamoto J.,
 RA Isono Y., Kawai-Hio Y., Saito K., Nishikawa T., Kimura K.,
 RA Yamashita H., Matsuo K., Nakamura Y., Sekine M., Kikuchi H., Kanda K.,
 RA Wagatsuma M., Murakawa K., Kanemori K., Takahashi-Fujii A., Oshima A.,
 RA Sugiyama A., Kawakami B., Suzuki Y., Sugano S., Nagahara K.,
 RA Masuho Y., Nagai K., Isogai T.;
 RL Submitted (JUL-2003) to the EMBL/GenBank/DBJ databases.
 DR EMBL; AK125816; BAC86305.1; -; mRNA.
 SQ SEQUENCE 166 AA; 17738 MW; ED8C12B80E6D5579 CRC64;
 Alignment Scores:
 Pred. No.: 7.77e-24 Length: 166
 Score: 301.50 Matches: 63
 Percent Similarity: 79.1% Conservative: 5
 Best Local Similarity: 73.3% Mismatches: 17
 Query Match: 32.5% Indels: 1
 DB: 2 Gaps: 1
 US-10-607-806-1_COPY_1_500 (1-500) x Q6ZUC0_HUMAN (1-166)
 QY 340 TCCAGGCTGGAGTGCAGTGGTGCAGTCTGGCTCACTGCACCTCCACCTCCACAGTTCA 281
 Db 61 SerArgLeuGluCysAsnGlyAlaIleSerAlaHisCysAsnLeuCysLeuProGlySer 80
 QY 280 AGTGATTTCTTCCGCCAGCCTCTCTGTAGCTGGGACTATAGCAAGTGCACCAACC 221
 Db 81 SerAspSerProAlaSerAlaSerGlnValAlaGlyThrThrGlyValCysHisAla 100
 QY 220 TGGCTAATT---TTTGTATTTTATTAGACATAGTGTTCCTGATGTGGCAGCTGGT 164
 Db 101 GlnLeuIleLeuPheValPheLeuValLysThrGlyPheCysHisValGlyGlnAlaArg 120
 QY 163 CTCAAACTCTGCTCTCAAGTGATCCACCACTCCGCTCTCCAAAGTGTGGGAAATACA 104
 Db 121 LeuGluLeuLeuThrSerSerAspProProThrSerAlaSerGlnSerAlaGlyIleThr 140
 QY 103 GACATGAGCCACCGCAC 86
 Db 141 GlyValSerHisArgThr 146
 RESULT 15
 ALU1_HUMAN
 ID ALU1_HUMAN STANDARD; PRT; 591 AA.
 AC P39188;
 DT 01-FEB-1995 (Rel. 31, Created)

01-FEB-1995 (Rel. 31, Last sequence update)
10-MAY-2005 (Rel. 47, Last annotation update)
Alu subfamily J sequence contamination warning entry.
Homo sapiens (Human)
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini; Hominidae;
Homo.
NCBI_TaxID=9606;
[1]
NUCLEOTIDE SEQUENCE
MEDLINE=95021758; PubMed=7935834; DOI=10.1038/371752a0;
Claverie J.-M., Makalowski W.;
"Alu alert."
Nature 371:752-752(1994).
[2]
CONCEPT.
MEDLINE=92241891; PubMed=1572661;
Claverie J.-M.;
"Identifying coding exons by similarity search: alu-derived and other
potentially misleading protein sequences.";
Genomics 12:838-841(1992).
[3]
ALU FAMILIES CLASSIFICATION.
MEDLINE=88333009; PubMed=3138422;
Quentin Y.;
"The Alu family developed through successive waves of fixation closely
connected with primate lineage history."
J. Mol. Evol. 27:194-202(1988).
[4]
ALU FAMILIES CLASSIFICATION.
MEDLINE=91178815; PubMed=1706781;
Jurka J., Milosavljevic A.;
"Reconstruction and analysis of human Alu genes."
J. Mol. Evol. 32:105-121(1991).
-!- MISCELLANEOUS: Various analyses indicate that Alu repeats fall
into 8 subfamilies (Ref.3 and Ref.4). Therefore, 8 Alu warning
consensus sequences have been constituted that contain all six
frames conceptual translations of each of these classes of Alu
repeats.
-!- MISCELLANEOUS: Isolated 'X' indicates the presence of a stop
codon, 'XXX' is used to separate the various translation phases.
-!- CAUTION: This Alu entry is provided in order to avoid the further
pollution of protein sequence databases with Alu-derived amino
acid sequences.
-!- CAUTION: Alu repetitive sequences are interspersed in human and
primate genomes with an average spacing of 4 kb. Some of them are
actively transcribed by pol III. Normal transcripts may contain
Alu-derived sequences in 5' or 3' untranslated regions. However,
cDNA libraries also contain partial and/or rearranged cDNAs
ligated with Alu-derived sequence in any orientation. Although Alu
elements (especially situated on the complementary strand) have a
great potential to create additional/alternative exons,
consideration should be given to the possibility that the presence
of an Alu in an open reading frame may have resulted from a
cloning artifact or may be due to misinterpretation of sequencing
data. This point has been overlooked on several occasions, with
the consequence of erroneous Alu-derived amino acid sequences
being reported.
-!- CAUTION: Any significant similarity of a putative protein sequence
with an Alu-translated entry must be taken as a warning that a
part of Alu repeat may have been artifactually included in the
coding nucleotide sequence.

This Swiss-Prot entry is copyright. It is produced through a collaboration
between the Swiss Institute of Bioinformatics and the EMBL outstation -
the European Bioinformatics Institute. There are no restrictions on its
use as long as its content is in no way modified and this statement is not
removed.

EMBL; U14567; -, NOT ANNOTATED_CDS; Genomic_DNA.
Hypothetical protein.
REGION 1 96 Frame-1.
REGION 100 195 Frame-2.

FT REGION 199 294 Frame-3.
FT REGION 298 393 Frame-4.
FT REGION 397 432 Frame-5.
FT REGION 496 591 Frame-6.
SQ SEQUENCE 591 AA; 63827 MW; 665D395735519D95 CRC64;
Alignment Scores:
Pred. No.: 1.18e-23 Length: 591
Score: 301.00 Matches: 56
Percent Similarity: 72.1% Conservative: 6
Best Local Similarity: 65.1% Mismatches: 24
Query Match: 32.4% Indels: 0
DB: 1 Gaps: 0
US-10-607-806-1_COPY_1_500 (1-500) x ALU1_HUMAN (1-591)
QY 338 CAGGCTGGAGTGCAGTGGTGCATCTTGGCTCACTGCACTCCAGTTCACG 279
DB 506 GlnAlaGlyValGlnTrpArgAspHisGlySerLeuGlnProArgProGlyLeuLys 525
QY 278 TGATTCTTCTGCCCCAGCTCTCTGTAGCTGGGACTATAGGCAAGTGCCACACACCTG 219
DB 526 ArgSerSerCysLeuSerLeuProSerSerTrpAspTrpArgAlaProProArgPro 545
QY 218 GCTAATTTTGTATTTTATTAGAGATAGTGTGTTTGGCATGTTGGCCAGGCTGCTCAA 159
DB 546 AlaAsnGlyCysLeuPheCysArgAspGlyValSerLeuCysCysProGlyTrpSerArg 565
QY 158 ACTCTCTCTCTCAAGTGATCCACCCCACTCGGCTCTCCAAAGTGCTGGGAATACAGACAT 99
DB 566 ThrProGlyLeuLys***SerSerArgLeuGlyLeuProLysCysTrpAspTrpArgArg 585
QY 98 GAGCCACCGCACCTGGCT 81
DB 586 GluProProArgProAla 591
Search completed: February 8, 2006, 09:43:15
Job time : 125.336 secs